

FIRST-NAMED INVENTOR OR
APPLICATION IDENTIFIER: Vinod Asundi et al.
Request for New Nonprovisional Application (37 C.F.R. §1.53)



Application No. (U.S.S.N.)	Filing Date (dd/mm/yy)	Status (Patented, Pending, Abandoned)
09/574,454	19/05/00	Pending
09/540,217	31/03/00	Pending
09/649,167	23/08/00	Pending
09/552,929	04/18/00	Pending
09/770,160	26/01/01	Pending
10/014,487	08/11/01	Pending
09/989,660	21/11/01	Pending
09/577,408	18/05/00	Pending

2. Specification (Total pages: 10966) consisting of:

Specification (10959 pages); Claims and Abstract (7 pages);
2 CD-ROMs (Tables 2-4, 6, 8-22);
(one original and one backup)
0 Sheets of Drawings;

In accordance with 37 CFR 1.58 (e)(4) Applicants state that copy 1 and copy 2 of the enclosed CD-ROM discs are identical.

3. Declaration and Power of Attorney (unexecuted, 7 pages);
4. Request to transfer sequence listing (2 pages);
5. Non-publication request (1 page);
6. Fee Calculation

TABLE 2: Nearest Neighbor (BlastN v. Genbank?)

SEQ ID NO:	ACCESSION	DESCRIPTION	SMITH-WATERMAN VALUE	% HOMOLOGU
1	AF026198	Fugu rubripes putative protein 2	677	62.195
2	AJ242724	Homo sapiens putative mitogen-activated protein kinase kinase kinase	1685	82.524
3	AL117567	Homo sapiens hypothetical protein	2042	99.385
4	U18237	Homo sapiens ATP-binding cassette protein	470	98.701
5	AL021366	Homo sapiens cICK0721Q.3 (Kinesin related protein)	4347	99.110
6	D00173	Homo sapiens cytochrome P-450	3275	100.000
7	X68061	Mus musculus HCNGP	1902	95.130
8	AF035940	Homo sapiens similar to mago nashi	968	100.000
9	AC005783	Homo sapiens R33083_1	2526	99.467
10	AB014599	Homo sapiens KIAA0699 protein	5341	100.000
11	AL050069	Homo sapiens hypothetical protein	1751	98.496
12	AF132963	Homo sapiens CGI-29 protein	1681	99.174
13	AF134983	Mus musculus energy-dependent regulator of proteolysis	3936	95.110
14	AF100753	Homo sapiens ancient ubiquitous 46 kDa protein AUP1	2748	100.000
15	AB015982	Homo sapiens serine/threonine kinase	6047	99.888
16	M64229	Homo sapiens type I collagen	181	96.429
17	Z81037	Caenorhabditis elegans predicted using Genefinder; Weak similarity in N-terminus to UNC-42 (WP:F58E6.1); cDNA EST EMBL:Z14323 comes from this gene	308	29.834
18	M35522	Canis familiaris GTP-binding protein (rab7)	449	49.624
19	U40952	Caenorhabditis elegans C03B1.10 gene product	200	65.000
20	AJ131395	Mus musculus collagen type XIV	128	34.211
21	AB023210	Homo sapiens KIAA0993 protein	2577	100.000
22	M64983	Homo sapiens fibrinogen beta chain	3373	99.793
23	D88010	Homo sapiens ribosomal protein S13	978	100.000
24	AF151848	Homo sapiens CGI-90 protein	433	34.211
25	L25314	Drosophila melanogaster actin-related protein	1269	47.328
26	M83104	Bos taurus cytochrome b-5 reductase	1278	64.260
27	AC005614	Homo sapiens F23269_2	4179	96.546
28	U90549	Homo sapiens non-histone chromosomal	601	100.000

1904	AJ011779	Homo sapiens SEC63 protein	4827	99.212
1905	X78933	Homo sapiens zinc finger protein	3469	99.590
1906	X01655	Homo sapiens type III procollagen (aa 892-1023)	170	34.400
1907	AF105332	Homo sapiens vitamin D3 receptor interacting protein	9085	99.345
1908	AF155113	Homo sapiens NY-REN-55 antigen	4902	96.370
1909	AL032821	Homo sapiens dJ55C23.1 (vanin 1)	3411	100.000
1910	X06562	Homo sapiens growth hormone receptor (AA 1-638)	4282	99.843
1911	M69181	Homo sapiens non-muscle myosin B	12358	99.545
1912	X61046	Hydra sp. mini-collagen	215	41.176
1913	X51798	Homo sapiens precursor	900	98.000
1914	AF084256	Homo sapiens beta glucuronidase isoform d	145	60.465
1915	AF152307	Homo sapiens protocadherin alpha 11	5459	89.053
1916	Y12711	Homo sapiens putative progesterone binding protein	939	90.955
1917	AB002330	Homo sapiens KIAA0332	6743	99.415
1918	AF083419	Homo sapiens calcium/calmodulin-dependent protein kinase II beta e subunit	3294	96.731
1919	AB020649	Homo sapiens KIAA0842 protein	6825	99.706
1920	AF123880	multiple sclerosis associated retrovirus element unknown protein U5/1	297	75.758
1921	AB002306	Homo sapiens KIAA0308	8346	99.692
1922	X06745	Homo sapiens DNA polymerase alpha-subunit (AA 1 - 1462)	9630	99.183
1923	AF028827	Homo sapiens Tax interaction protein 40	1520	99.567
1924	X71428	Homo sapiens FUS glycine rich protein	3601	95.336
1925	D86984	Homo sapiens similar to yeast adenylate cyclase (S56776)	3034	99.580
1926	AF093250	Homo sapiens P38IP	4708	99.864
1927	AF151877	Homo sapiens CGI-119 protein	1545	99.598
1928	AF165124	Homo sapiens gamma-aminobutyric acid A receptor gamma 2	3150	99.786

TABLE 3: Nearest neighbor (FastA v. Genbank, Genpept release 114)

SEQ ID NO.	ACCESSION NO.	DESCRIPTION	SMITH- WATERMAN SCORE	% IDENTITY
1	L27428	Homo sapiens reverse transcriptase	253	32.129
2	X97675	Homo sapiens plakophilin 2b	167	50.794
3	U49082	Homo sapiens transporter protein	1001	56.122
4	U17247	Saccharomyces cerevisiae Imhlp	164	25.217
5	Z38061	Saccharomyces cerevisiae mal5, stal, len: 1367, CAI: 0.3, AMYH YEAST P08640 GLUCOAMYLASE S1 (EC 3.2.1.3)	320	26.144
6	AF080234	Human endogenous retrovirus K polymerase	330	46.774
7	M13101	Rattus norvegicus unknown protein	265	49.398
8	M12140	Homo sapiens envelope protein	458	40.865
9	U49974	Homo sapiens mariner transposase	545	78.571
10	AB014549	Homo sapiens KIAA0649 protein	588	81.818
11	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	266	65.789
12	AB007903	Homo sapiens KIAA0443	1527	65.445
13	X78933	Homo sapiens zinc finger protein	465	65.979
14	AF081111	Mus musculus domesticus ORF2	276	63.158
15	AF081111	Mus musculus domesticus ORF2	178	46.341
16	AB012223	Canis familiaris ORF2	271	49.485
17	X03725	Mus musculus ORF 2 (466 aa)	222	43.902
18	U49974	Homo sapiens mariner transposase	870	71.491
19	D49677	Homo sapiens U2AF1-RS2	476	67.391
20	AJ001714	Homo sapiens Myosin-IXA	261	58.750
21	U93572	Homo sapiens putative p150	295	50.000
22	AB012223	Canis familiaris ORF2	182	43.434
23	X53581	Rattus norvegicus ORF4	176	40.230
24	L27428	Homo sapiens reverse transcriptase	223	35.714
25	U93568	Homo sapiens putative p150	227	65.574
26	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	285	52.542
27	AF003535	Homo sapiens ORF2-like protein	252	42.149
28	AF123881	multiple sclerosis associated retrovirus element gag polyprotein	549	84.821
29	AF123881	multiple sclerosis associated retrovirus element gag polyprotein	391	62.500
30	AB022046	Cynops pyrrhogaster alpha type II collagen	151	35.294
31	AF015539	Mytilus edulis precollagen P	265	32.950
32	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	367	57.983
33	AL032660	Caenorhabditis elegans predicted using Genefinder	930	37.174
34	AJ005073	Mus musculus Alix	2666	88.565
35	AL117237	Homo sapiens hypothetical protein	523	72.656
36	U93570	Homo sapiens p40	190	33.333
37	U09116	Homo sapiens ORF2, encodes a reverse transcriptase homolog	248	47.107
38	M34651	Pseudorabies virus ORF-3 protein	159	40.206
39	L76559	Drosophila melanogaster mus308	271	51.724
40	X06745	Homo sapiens DNA polymerase alpha- subunit (AA 1 - 1462)	382	68.317

10284	AF009329	Rattus norvegicus enhancer-of-split and hairy-related protein 1	529	98.718
10285	AL049558	Schizosaccharomyces pombe hypothetical protein	267	37.725
10286	L15309	Homo sapiens zinc finger protein	237	85.714
10287	Z98596	Schizosaccharomyces pombe putative SMC family protein	302	55.000
10288	U13152	Mesocricetus auratus guanine nucleotide-binding protein beta 5	362	53.636
10289	X93302	Saccharomyces cerevisiae Msn5 protein	173	35.200

TABLE 4

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	12 90 158 166 185 187-188 190 195 208-209 225 227 240 260 270 315 338 396 400-401 457 594 640-641 654 692 714 740 786 835 849 852 867 905 942 983 1066 1076 1087 1102 1107 1289 1302 1317 1464 1492 1679 1749 1822-1823 1846 1927 1945 2039 2164 2217 2372 2384 2448 2818 2820 2823 2834-2835 2862 2877-2878 2884 2891 2896 2906 2914 2918-2919 2932 2950 2967 2973 2978 2983 2997 3003 3010 3012-3013 3045 3066 3068 3091 3094 3110 3130-3131 3135 3165 3183 3213- 3214 3223 3240 3247 3257 3260 3265 3276 3284 3312 3327 3331 3381 3395 3411 3427 3441 3448 3453 3461 3465 3469 3471 3492 3499 3506 3508 3515 3534 3536 3542-3543 3559 3566 3613 3625 3631-3632 3638 3646 3655 3663 3666 3668 3683 3686 3693 3740 3745 3747-3748 3758 3762 3779 3782-3783 3866 3871-3872 3877 3919 3943-3944 3965 3978 3994 3996 4000 4006 4008 4010 4036 4040 4043 4057 4060 4068 4074 4093 4097 4106 4122 4139-4140 4143 4145 4150 4178 4193 4228 4243 4247 4252 4254 4258 4264 4274 4287 4300 4303 4310 4316-4317 4322 4324 4343 4360 4379- 4380 4391 4403 4407 4409-4410 4421 4423 4426 4429 4432-4435 4437 4447 4455 4464 4472 4481 4488 4492 4494 4499 4501 4503 4506 4518 4526- 4529 4532 4541 4543 4545 4547-4548 4552 4569 4571 4580 4585 4587 4632 4642 4646 4654 4659 4665 4683 4706 4715 4734 4758 4773 4775 4778 4787 4797-4798 4811 4820 4852 4868 4873 4878 4888 4899 4912 4917-4918 4929-4930 4935 4940 4944 4947 4968 4971 4989 4998 5026 5063 5096 5098 5112 5115 5120-5121 5124 5126 5132 5136 5163 5171 5177 5189 5196 5199 5216 5235 5259 5279 5289 5292 5294 5298 5376 5382 5403 5415 5443 5450 5459 5462 5469 5476 5540 5551 5554 5563 5582-5583 5589 5596 5612 5618 5622 5624 5627 5645 5651 5663-5667 5681 5698 5700 5703 5718 5720 5733-5735 5764 5775 5787-5788 5791 5796 5800 5804 5824-5825 5832 5842 5863 5868 5897 5908 5921 5929 5950 5961 5974 6032 6039- 6040 6051 6057 6067 6069 6078 6096 6100 6105 6114 6131 6134 6143 6145 6153 6157 6168 6176 6188 6192 6204 6211 6216 6222 6226-6230 6232- 6233 6252 6263 6280-6281 6285 6292 6296 6313 6321 6323 6325 6327-6328 6330 6332-6334 6336- 6337 6343 6355-6356 6384 6391-6392 6407 6414 6416-6417 6422 6448 6456 6459 6480 6491 6502 6507-6508 6512 6517 6522 6525 6534 6538 6541 6547-6548 6553 6562-6563 6572 6575 6610 6618 6621 6624 6629 6642 6671 6684 6690 6692 6714 6730 6742-6744 6757 6787 6790 6797 6804-6805

			8827 8876 8907 8917 8952 8963 9031-9032 9075
			9090 9119 9151 9175 9205 9274 9349 9550 9590
			9600 9667 9747 9839 9937 10164

TABLE 6: Nearest neighbor (FastA v. Genbank, Genpept release 114)

SEQ ID NO.	ACCESSION NO.	DESCRIPTION	SMITH- WATERMAN SCORE	% IDENTITY
10	D64006	Synechocystis sp. asparaginyl-tRNA synthetase	997	42.359
11	Z74036	Caenorhabditis elegans predicted using Genefinder; similar to collagen	176	32.847
12	AB023195	Homo sapiens KIAA0978 protein	117	29.008
13	U49973	Homo sapiens ORF1; MER37; putative transposase similar to pogo element	254	62.069
14	AF053091	Drosophila melanogaster eyelid	189	23.211
15	Z28082	Saccharomyces cerevisiae ORF YKL083w	111	28.182
16	U93571	Homo sapiens p40	245	37.398
17	AC004602	Homo sapiens F23487.2	405	59.259
18	U70935	Peromyscus maniculatus reverse transcriptase	323	50.847
19	M22334	Homo sapiens unknown protein	306	50.909
20	U29156	Mus musculus involved in signaling by the epidermal growth factor receptor; Method: conceptual translation supplied by author	276	50.000
21	A27950	Ovine pulmonary adenocarcinoma virus gag	224	32.215
22	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	355	47.486
23	U20979	Homo sapiens chromatin assembly factor-I p150 subunit	228	62.195
24	D00570	Mus musculus open reading frame (196 AA)	112	42.647
25	L34685	Arabidopsis thaliana cell wall protein	91	29.487
26	AE001164	Borrelia burgdorferi exodeoxyribonuclease V, beta chain (recB)	50	26.786
27	M90464	Homo sapiens alpha-5 type IV collagen	124	34.307
28	U20286	Rattus norvegicus lamina associated polypeptide 1C	835	58.436
29	AB016092	Homo sapiens RNA binding protein	134	31.000
30	AJ243460	Leishmania major proteophosphoglycan	255	34.016
31	X61296	Rattus norvegicus open reading frame 2	179	38.281
32	M22334	Homo sapiens unknown protein	219	57.303
33	AL033534	Schizosaccharomyces pombe serine-rich protein	124	30.709
34	U93570	Homo sapiens putative p150	348	56.557
35	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	283	31.178
36	Y00706	Homo sapiens alpha1-chain	93	33.588
37	AL049801	Homo sapiens hypothetical protein	283	29.536
38	U49973	Homo sapiens ORF1; MER37; putative transposase similar to pogo element	608	67.358
39	M13002	Mus musculus 2855 is the position of the first start codon in ORF 2; putative	277	35.338
40	AB017600	Haliotis discus collagen pro alpha-chain	188	29.940
41	X89430	Homo sapiens methyl CpG binding protein 2	232	55.128
42	AF003535	Homo sapiens ORF2-like protein	218	50.000
43	L14850	Microciconia prolifera collagen	163	30.804
44	M91463	Homo sapiens glucose transporter	419	59.028
45	Y07687	Mus musculus nuclear factor i	103	24.855
46	M22334	Homo sapiens unknown protein	166	29.150
47	X89453	Rattus norvegicus DRPLA	200	25.130
48	X56015	Crithidia oncopelti NADH dehydrogenase subunit 5	106	30.108
49	AF131877	Streptomyces collinus NapG oxidoreductase	151	31.447
50	AF159461	Mus musculus RFG	450	64.238

3775	AE001793	Thermotoga maritima hypothetical protein	56	58.333
3776	M80344	Homo sapiens ORF1 codes for a 40 kDa product	332	43.939
3777	AL117203	Caenorhabditis elegans predicted using Genefinder; preliminary prediction	82	45.161
3778	AE001098	Archaeoglobus fulgidus A. fulgidus predicted coding region AF0123	87	37.500
3779	AF085753	Erimystax dissimilis NADH ubiquinone oxidoreductase subunit 2	84	32.609
3780	U60669	Homo sapiens 1,25-dihydroxyvitamin D3 24- hydroxylase	95	38.889
3781	U60669	Homo sapiens 1,25-dihydroxyvitamin D3 24- hydroxylase	99	39.189
3782	M18327	vectors LacOPZ-alpha peptide from pUC9; putative	106	56.000
3783	U15370	Pseudomonas aeruginosa similar to E. coli protein PolB/DinA, Swiss-Prot Accession Number P21189	79	45.000
3784	AB001684	Chlorella vulgaris ORF54d	82	27.500
3785	Z72511	Unknown possible zinc finger protein; cDNA EST EMBL:M89115 comes from this gene; cDNA EST EMBL:D715	1343	57.576
3786	U09453	Cricetulus griseus UDP-N-acetylglucosamine: dolichyl phosphate N-acetylglucosamine 1-phosphate transferase	234	66.667
3787	Z93386	Unknown Similarity to Yeast hypothetical 52.9 KD protein (SW:P43616); cDNA EST EMBL:M89432 comes fr	180	38.889
3788	X58906	Homo sapiens steroid 21-monooxygenase	785	91.270
3789	AL050015	Homo sapiens hypothetical protein	606	87.379
3790	U19729	Saccharomyces cerevisiae Ylr404wp	458	100.000
3791	Z81489	Homo sapiens Similarity to Human host cell factor C1 (SW:P51610); cDNA EST yk205a3.5 comes from this gen	1055	40.000
3792	U80448	Unknown coded for by C. elegans cDNA CEES55F; coded for by C. elegans cDNA yk84a1.3; coded for by	1112	54.493
3793	Z37166	Homo sapiens nuclear RNA helicase (DEAD family)	2817	100.000
3794	X04494	Homo sapiens precursor polypeptide	2437	100.000
3795	AF015811	Mus musculus putative lysophosphatidic acid acyltransferase	1900	96.564
3796	AJ007395	Homo sapiens QA79 membrane protein	3166	100.000

TABLE 8

SEQ ID NO.	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of amino acid sequence	Predicted end nucleotide location corresponding to last amino acid residue of amino acid sequence	AMINO ACID SEQUENCE ENCODED BY THE NUCLEOTIDE SEQUENCE OF THE CORRESPONDING SEQ ID NO. (A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=possible nucleotide deletion, \=possible nucleotide insertion)
1	C	31	162	MPGPFGSLYLLLWNIPVTSSNSFHSWQGNFLKFTRNTTKRNR*
2	A	337	1	TPSLIHQAPTPCPAGLWG/PPNGHYHGS*PGCHWPQAPHRA***GLLPPRWL GHGLPGGPAAPWAASQWVDGVAGRLPGPAWSWHASGAAPAQPGPL*LLVPGS SGLPDPDP
3	A	1	379	NFKHVCLFTRPAGASLTAP*PGNAI*PGGPREGLSMPGRSPGQGYLRAAHL KGQCLRPSTPSPSPGDPGHAGCQP\G\RPLCQDET EGAPPPGNGSSGVI RTGQKQIAGLGVI INVGTSFAVS
4	A	29	317	SSSYNSSSTFIYVCLFCAQSIILR/CLFSLSKYIGSMLFITNLRLLYLLHLSI LNH*CHMYCIPKFTI*NTWLFRCRVFLFISFFELKLKENQGPKEG
5	A	2	468	RLFQILVQIYSIFSLFSAYLFIVQYSLNLLFLMIHF*QVDLTDVSKFHSYQ RLIYVYYY*LDETMSNK\DSRNEYTSPQSORE*T*STL*FVFSVLLQGYS* PTD\SYKIIIFIKLT*CSVA*I*KIDKIMYNKTLIKLNRYCNIEQFFITSF Y
6	A	13	534	TSAPVLSPSPSLPSSLTSPPLPLPLPLPLPLVHPPSLLLPLNYPSPASP IRLRSHPSQAYPPPPVLPTGQPTCLLYLASL*FPASLPLGPSVFPLAFVHF LPLHPLSPLSLAALSPPXRRTPPPLSPLLPCCQALDLAAXXLVRCGLDLPQ TSLHFQQQTPSTHDPWPL
7	A	3	397	STGLAYIILTETLKLKSLSTQ\FLT KPLCPDQLQTRAELSMELSNKWDQDPE GLL*RTVTGDETGLFQYDPENKAQ/SKPRGGSGPVKANVDQSRAEVMAMVFW NAQGILLVDFLKSRPITSAYYESIYYEK
8	C	47	202	MIKTQCYTFIPNSTAPNKSITKALQGLTALSNEEDSNGVNDPFTGWLKKK*
9	A	3	319	SLELLVFSLYGSWCKGDSLLEEKQKLIRPEHRRMNLAAPCGRHCVFLLWDH WEGIKGSRVGLVWSGQV/VKEGTHQWKR DG*LCLNAKMFLCLSILFKIHKC LC
10	A	667	1	VAPASPMGPL\DLSSGPQSGRSPEAPGPSTLSSSGQFSAWHTKDPDTSRPR HTPSKDPIGAHPGSITQNPSSGFSDLHPTFEYPSIPQFPNEYSHPNTEVPS LMGSPSPDPHSPSPQSSSKM*PLPLCVFSPSPVSPGYGHSVFQPGRLPLLA IPTHSFLPGSVSGLLPPIRQYPMWDSGGSGIPAFHLFFHLFVLTCTKKR GLYYHRLFLNRR
11	A	74	651	GWLCAGWGGVPGSVSGGSSSTSFVINHCRGSACTKCGPGRADAAAG*GGLTD PFPQKR\PPKGIAPEEGSPRPH*KGTP/QGHPPKEGSSKGPPNGVPPRLPS EGGPRRLPP*RCPPRPAPWKRAQ/GPLPLEEPPKFTLEKGLSQSFPGRAP SSWAGREELCTRHP*DLLSSRFPCPGPMNPEAVLEPDG
12	A	1	392	GTRALCAGACLHAHLEMLNR*THLDPESKGRFVPHSHDCPLWPSEPLRD \VMCTAFHDNEETFLT KYLYEIATRHSYFYAPELLFFAKRDPFAFTECCQAA D*AACLLPKLDELRYEGKASSAIQRLK

				AFQKVADMATNTRGGTREIGSALTRMCMRHSIEAKLRQFSSALIDCLINPL QEQMEEWKKVANQLDKDHAKEYKKARQEIKKKSSDTLKLQKKAKKGRGDIQP QLDSALQDVNDKYLLLEETEKQAVRKALIEERGRFCTFISMLRPVIEEEISM LGEITHLQTI SEDLKSLTMDPHKLPSSSEQVILDLKGS DYSWSYQT PPSSPS TTMSRKSSVCSSLNSVNSSDSRSSGSHSHSPSSHYRYSNLAQQAPVRLSS VSSHDSGFISQDAFQSKSPSPMPPEAPNQRRKEKREPD PNGGGPTTASGPPA AAEAAQRPRSM
10410	A	511	958	AGRGGPGKPVSWSSGPGSPGQTQRRSWVKSTRGHSSLLPPSQDFVAGLSVIL RGTVDDRLNNAFNLYDLNKDGCITKEEMLDIMKSIYDMMGKYTYPALREEAP REHVESFFQKMDRNKDGVTIEEFIESCQKDENIMRSMQLFDNVI

TABLE 9

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	18-19 49 87-90 240 246 424 471 835 944 1004 1021 1062-1063 1066 1092 1461 1538 2012-2015 2227 2260 2277 2326 2583 2669 2942 2960 3107 3142 3176 3193 3196-3201 3242-3244 3250 3363 3398 3403 3420 3428 3431 3439 3490 3505 3530 3571 3753 3814 3822 3854 3956 3994 4020 4065 4101 4115 4159 4167 4196 4228-4229 4236 4238 4246 4259 4263 4267 4296 4338 4356 4380 4382 4407 4470 4494 4502 4521 4529 4549 4594 4599 4647 4653 4661 4679 4681 4705 4744 4769-4770 4805 4867 4869 4900 4902-4903 4913 4928 4993 5073 5154 5208 5245 5249 5261 5268 5301 5331 5347 5351-5358 5363 5375 5399 5408 5478 5505 5572 5621 5640 5664 5685 5701-5702 5705 5714 5731 5736 5741 5757-5758 5763 5789 5805 5859 5951 5992 6007 6037 6056 6061 6078 6103 6116 6119 6147 6156-6157 6159 6171 6183 6187 6195- 6196 6200 6203 6215-6216 6221-6222 6224 6226 6234 6239 6245 6248 6269 6287 6294 6297 6344 6356 6362-6363 6378 6380 6394 6419 6425 6455 6457 6467 6534 6538 6540 6557 6559 6562 6585 6590 6593 6595 6601-6604 6609-6610 6616 6628 6630 6635 6646 6659 6663 6687 6690 6713 6718 6732 6734 6739 6752 6756-6757 6765 6770 6776 6785 6810 6842 6844 6854 6868 6880 6888 6895 6900 6909 6918 6921 6932 6956-6957 6960 7008 7030 7034 7041 7045 7064 7073 7088-7089 7092 7111 7125 7136 7163-7165 7172 7211 7214 7228 7310 7335 7365 7379 7405 7454 7472 7539 7575 7590 7599 7605 7608 7613 7618 7630 7637 7639 7642 7644 7646-7647 7649 7654 7674 7678 7692 7699 7711 7719 7721 7726-7727 7731-7736 7738 7745 7747 7750-7751 7766 7770 7773 7776-7778 7789 7795 7802-7805 7807 7818 7826 7829 7832 7864 7878 7884 7896 7915 7924-7925 7943-7944 7948 7967 8000 8004 8007 8009 8012 8018 8022 8045 8064 8093 8103 8111 8123 8133 8139 8164 8167 8172 8174 8181 8186 8188 8201 8213 8215- 8216 8219 8224 8233 8241 8259 8261 8273 8281 8289-8290 8293-8294 8302 8305 8369 8380 8405 8408 8411 8427 8434 8448 8451 8461 8510 8525 8528 8545 8547 8560 8565 8580 8584 8587 8616 8643 8671 8683 8777 8795 8815 8820 8848 8857 8870 8881 8885 8893-8894 8923 8930 8967 8988 9077 9142 9188 9201 9224 9226 9260 9420 9431 9505 9546 9568 9599 9665 9733-9734 9745 9811 9866 9875 9952 10021 10260 10272 10339 10344 10350 10352 10356 10358 10367 10384 10392 10398 10404 10407

TABLE 9

			9537 9554-9555 9834 9838 9902 9941 9952 10003 10104 10125 10133 10167 10235 10339-10340 10348 10358
uterus	Clontech	UTR001	134 316 456-464 597-598 663 992 1092 1176 1327 1501 1576 1638-1640 1705 2227 2235-2236 2351 2382-2383 2582 2779 2799 3043 3136 3190 3198 3232 3339 3420 3548 3726 3758 3793 3822 3848 3994 4139-4140 4202 4206 4300 4311 4329 4345 4360 4388 4405 4409-4410 4470 4497 4500 4520 4555 4601 4661 4672 4674 4681 4723 4784 4858 4867 4903 4928 4953-4954 4956 4994 5031 5073 5155 5207-5208 5244 5421 5430 5445 5539 5544 5605-5606 5614 5675 5685 5732 5741 5768 5784 5805 5886 5972 5998 6083 6119 6147-6148 6167 6187 6201 6210 6232 6239 6287 6305 6312 6336- 6339 6354 6360 6369 6394 6421 6457 6464 6494 6503-6504 6510 6532 6534 6544 6549 6551 6553 6559 6599 6604 6616 6622-6623 6630 6659 6687 6709 6714 6727 6739 6765 6769 6777 6791 6794- 6797 6816 6819 6842 6850 6865 6909 6930 6948 6958 7002 7013 7035 7084 7127 7134 7327 7546 7575 7579 7590 7592 7605 7608 7618 7625 7641 7645 7652 7664 7671 7673 7686 7694 7715 7719 7721 7743 7758 7782 7784 7818 7822 7884 7943 7948 7953 7974 7978 7980 7998 8004 8007 8019 8045 8064 8099 8133-8136 8163 8165 8172 8190 8197 8214 8228 8268 8273 8318 8338 8342 8346 8358 8387 8411-8412 8427 8435 8451 8461 8534 8564 8572 8580 8608 8616 8640 8654 8660 8752 8772-8773 8818 8839 8844 8853 8855 8866 8897 8959 8983 8993-8994 9077 9163 9194 9310 9324 9476 9510 9521 9596 9665 9710 9891 9903 10127 10131 10167 10344 10358 10375 10407

Table 10: Nearest neighbor (FastA v. Genbank, Genpept release 114)

SEQ ID NO.	ACCESSION NO.	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	M55261	Blastochloris viridis light harvesting complex alpha-subunit	70	34.043
2	X13783	Homo sapiens alpha-1 type 2 collagen (714 AA)	99	36.190
3	Z81503	Caenorhabditis elegans predicted using Genefinder; similar to collagen; cDNA EST EMBL:D65450 comes from this gene; cDNA EST EMBL:D68888 comes from this gene	140	38.710
4	U38184	Trypanosoma cruzi ATPase subunit 6	95	37.333
5	AE000323	Escherichia coli putative ligase	108	23.944
6	AL022537	Arabidopsis thaliana putative protein	256	35.366
7	U49974	Arabidopsis thaliana putative protein	585	76.154
8	AF038599	Homo sapiens mariner transposase	114	30.612
9	AB010886	Sus scrofa env protein	116	46.341
10	AC007202	Cydia pomonella granulovirus chitinase	198	30.000
11	L17318	Arabidopsis thaliana T8K14.10	233	35.673
12	AF190168	Rattus norvegicus proline-rich proteoglycan	512	68.254
13	AF010144	Homo sapiens serum albumin precursor	144	46.535
14	L76559	Homo sapiens neuronal thread protein AD7c-NTP	301	55.172
15	AB012223	Drosophila melanogaster mus308	377	52.830
16	X03725	Canis familiaris ORF2	291	51.064
17	Y12713	Mus musculus ORF 2 (466 aa)	147	48.077
18	M22432	Mus musculus Pro-Pol-dUTPase polyprotein	391	70.000
19	AF032103	Mus musculus elongation factor Tu	237	47.778
20	U51723	Homo sapiens ataxin-7	396	33.514
21	AB023203	Plasmodium vivax V-SERA 4	756	42.105
22	AF029676	Homo sapiens KIAA0986 protein	34	31.818
23	M64983	Bactrocera tryoni doublesex	292	73.846
24	M13101	Homo sapiens fibrinogen beta chain	241	38.136
25	AB023482	Rattus norvegicus unknown protein	180	41.975
26	X53581	Oryza sativa Similar to Zea mays PRP gene. (X60432)	202	46.575
27	AB012223	Rattus norvegicus ORF4	247	49.412
28	X52235	Canis familiaris ORF2	275	50.000
29	Y12713	Homo sapiens ORFII	359	60.000
30	AC003027	Mus musculus Pro-Pol-dUTPase polyprotein	497	40.343
31	V01201	Arabidopsis thaliana lcl prt_seq No definition line found	173	39.583
32	X17025	Rattus norvegicus coding sequence of pol	439	66.038
33	AL117200	Homo sapiens homologue of yeast IPP isomerase	180	27.381
34	X69490	Caenorhabditis elegans predicted using Genefinder; preliminary prediction	397	82.558
35	AB000450	Homo sapiens titin	530	70.732
36	M13100	Homo sapiens VRK2	187	35.780
37	AB014548	Rattus norvegicus unknown protein	483	68.067
38	U03279	Homo sapiens KIAA0648 protein	331	50.000
39	AF044321	Mus musculus phosphatidylinositol 3-kinase 110 kDa subunit	171	72.727
40	D12981	Homo sapiens cytochrome c oxidase assembly protein COX11	391	72.222
		Homo sapiens paraneoplastic cerebellar		

		imprinted gene, Peg3.		
10393	L35604	Drosophila melanogaster ethanolamine kinase	921	40.921
10394	U54807	Rattus norvegicus GTP-binding protein	1483	97.357
10395	AB002359	Homo sapiens KIAA0361	9350	99.927
10396	Z81564	Unknown cDNA EST yk448c11.3 comes from this gene; cDNA EST EMBL:C09223 comes from this gene; cDNA E	289	32.609
10397	AL109630	Drosophila melanogaster BACR7A4.h	737	32.083
10398	D10376	Bos taurus mitochondrial adenylate kinase isozyme 3	1405	92.511
10399	A01771	synthetic construct vascular anticoagulating protein	2092	99.694
10400	Z99162	Schizosaccharomyces pombe hypothetical protein	137	26.316
10401	AF131851	Homo sapiens Unknown	325	49.438
10402	M55542	Homo sapiens guanylate binding protein isoform I	3356	87.372
10403	AF026528	Rattus norvegicus stathmin-like-protein RB3	1215	99.471
10404	AL021106	Unknown /prediction=(method:"genefinder", version:"084"); /prediction=(method:"genscan", ve	751	50.917
10405	AF132157	Drosophila melanogaster unknown	115	23.611
10406	AE000717	Aquifex aeolicus hypothetical protein	175	26.277
10407	Z48334	Caenorhabditis elegans F10B5.8	2274	68.113
10408	AC006955	Arabidopsis thaliana unknown protein	217	34.615
10409	U70669	Homo sapiens Fas-ligand associated factor 3	226	23.894
10410	AF120102	Homo sapiens calsenilin	589	78.704

TABLE 11

SEQ ID NO.	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of amino acid sequence	Predicted end nucleotide location corresponding to last amino acid residue of amino acid sequence	AMINO ACID SEQUENCE ENCODED BY THE NUCLEOTIDE SEQUENCE OF THE CORRESPONDING SEQ ID NO. (A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=possible nucleotide deletion, \=possible nucleotide insertion)
1	A	114	436	AIFKCVEGMFRIAMVNVCFVSSGSLLI*PLTY/GVYDEWTHFAYM TIDLLEIPITGSHPVVLNALFCLEAP\WISPNTGSFAYPVYPKSL IAHDFAVEATMPYIRLSST
2	A	124	467	KSNIPNLGDCGWESLFNR\QSWRSSLAV\NDTYSSKKSNAETFTF HADLCTLSKDRPITIQTALAEVLVHKHPKATYEQLIAVLDEF/AN FLKKWWKAYDKENLFCEEG*KLCAASN
3	A	2	453	SKQMITIHNTKGR*SLILVSLIIFIATTNLLGLLPYSFTPTTQL SINLAMAIP*AGAEVICLQFFFN/NSLAS*/MPKVRIITMIIVML SIYRIISFSDLSTILFISVLLSITRVCVYIYSSPHVVFVYKCV CSIPSTRGYPFVLRSDL
4	A	1	427	EGFLELLRTRNHSNSQLQLTTGIGLFLNEGLKLVDFLEDV*K*Y HSETFTVNFSDTE*AMKHINDYVEKGTQGIKVDLVKELDRDVTVD LANYIFFKGKWDPRPFVNDTEEDFHVDQVSTVNEPIMKLLS\ML NIHPCFKL
5	A	1	465	KIKSFYASKDTIKMRVTDWKKIFAY*ISDKELIFTLRLKT**K *GKQPNLKNQGEI*VPISPQDIQIAHK\HLEGWSTSLVSELAPCE APV/RHPLTGLTIAGLQGFGEAGRLVRGRWGC*W/VHPF*KYI/W QFLSKLHISLPYDPTTPLLGTCSR
6	A	308	3	HFVIHSHKDLAIAHLGIY/PREMKT*VHTKTCT*IFTVALSVIAR AWNQPGRPLCSEWL\KYMVHTME*HSAIKRLNRYKNNCVNLFGL ITLSEKSQTQNV
7	A	587	2	FLTRETGDPTGRSSSHANTQSRFFPDPPG\PLNNLGNTHGCGRR AGRCPGTGPDPG\AGCGGPRCWPSGHLAATGD*GPSCGRLGANRG EAGPAGFTACSPLSGCRTPYTHFPASRMSCHLNCASPRTYRSQG NRGCERVAQGSQAGGERGAKSQVPVPAPARNKDPACRCKPRNRR PGNSGPVVRAYRRQR
8	A	1	475	RILNEEHGKYEGLHE*EVKWHLYIKSPAFTDLHLCYQKDMNGIST SASSPAVGTVGMDMEDDDFSKWNFYSPHSYPDK*LTIFKTESR VRESDEVTQIKVNWDEEVISGLLTSKDNVLKATGVLYDYAYK\Y LCEHTRSTLKEESLKLERNLQNH
9	A	3	539	SQCSPFISPACSLTALEEETEALRVHPRCLSPNLAPSSGPPRPP ELAPCPPSSQAGLRTCHSWVKGLHQPLPVASGMKSTFCNKTYTCP YPPP/PPLCPNHSNALTLPDSVTHAVPFE/L*SPSAPPSSTA*I LGSPS\CGASPCNHPHSHPGICPTPPGLWPVCPAPRAWQRDGR QT
10	A	2	453	RL*LGLEYALLVWGTPKV*H*GGFPYIYIVLLLSYALHQVTEYS MYVSIMAFNAKVS DPLIVGTYMTLLNTVSNLIGNWPSTVSLWLNV PLTVKECV*TSYQNCCTPDAAELCKKLGGSCVTALDGYVVEIIC VSIAFV/W*VFLVHKFK
11	A	20	475	KMGVPLLMSDPNRFLLFPKNFLREKTISPPKTF*PLKIWVKQWV LNFLGFPFGFKIIFPVFKFFFFFFF/RDRVSLYHPGWSAVSQSEL

TABLE 11

				KTDEPKGKTEAYLEAIRKNIEWLKKHDKKGNKEDYDLSKMRDFIN KQADAYVEKGILDKEEAIAIKRIYSSL
14069	A	1	140	NTSVAIQTIEYYSAFKRKETLTHVPLWMNLKDTMLREISQSQKDT V
14070	A	3	1654	QNAKVGSEAIQVLPAPFPREVQRLPEPAKSCRQRQLEARKAAAEKK KIQKEKLSTPEKIQEAELELAGITSDPGLSLKGGLSQQGLKPSLK VEPQNHFSFKEYSGNAVVESSVGLNCRPSDPYSMNSVYSYHSYY AQPSLTSVNGFHSKYALPSFSYYGFPSSNPVFPSSQFLGPGAWGHS GSSGSFEKKPDLHALHNSLSPAYGGAFAELPSQAVPTDAHHTP HHQQPAYPGPKYLLPKAPLLHSVSRDPSPPAQSSNCSNRSIKQE PVDPLTQAEFVPRDAGKMGKTPLEVSQNGGPSHLWGQYSGGSPM SPKRTNGVGGSWGVSFSSGESPAIVPDKLSSFGASCLAPSHFTDGO WGLFPGEGQQAASHSGGRLRGKPWSPCKFGNSTSALAGPSLTEKP WALGAGDFNSALKGSPGFQDKLWNP MKGEEGRIPAAGASQLVFYQ HKNLNQPNHGLALWEAKMKQLAERARARQEEAARLGLGQQEAKLY GKKRKWGGTVVAEPQQKEKKGVVPTRQALAVPTDSAVTVSSYAYT KVTGPYSRWI
14071	A	212	104	HTNHQCQTVLPAGQALATPQGLAPSPVFPQSCLRMV
14072	A	358	112	FPLTVMLNVFLWWQNLKHFCWFIEQQLMFFFFFFLRTGSHYVVQA GLELLGSTNLPAVSRSVAGTTGIHNCTQFNPLHTLHL
14073	A	3	151	YMGFHHVGYAGLELLTSSDLPPWPPKCWDYRHEPSCLAMFFYFAL IAPE
14074	A	8	1493	VTIHHLFVQGAVRADTLKKKSPSCLLCIVFIPVPRIDL TNIDQVA VIFKHHPVGRGDAVLKTWAPAQCLCSRMPAWLWLLGTGILASV HCQPLLAHGDKSLQGPQPPRHQLSEPAPAYHRITPTITNFALRLY KELAADAPGNIFFSPVSISTTLALLSLGAQANTSALILEGLGFNL TETPEADIHQGFRSLHTLALPSPKLELVGNSLFLDKRLKPRQH YLDSEKELYGAFAFSANFTDSVTTGRQINDYLRRQTYGQVVDCLP EFSQDTFMVLANYIFFKAKWKHPFSRYQTQKQESFFVDERTSLQV PMMHQKEMHRFLYDQDLACTVLQIEYRGNALALLVLPDPGKMKQV EAALQPQTLRKWGQLLLPSLLDLHLPRFSISGTYNLEDILPQIGL TNILNLEADFGVTGQLNKTISKVSHKAMVDMSEKGEAGAASGL LSQPPSLNTMSDPHAHFNRPFLLLLWEVTTQSLFLGKVNPVAG

TABLE 12

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	88 550 603-604 606-613 622 848 1011 1155 1354 1501 1988 2153 2156 2158 2176 2281 2362 2380 2528 2586 2895 2899 3022 3121- 3123 3127-3128 3132-3133 3137-3139 3221 3295 3305 3307 3379 3448 3560 3657 3712 3732 3820 3833 3862 3922 4030 4045-4046 4049 4067 4218 4225 4277 4301 4322 4417 4428 4439 4698 4702 4753 4778-4779 4781 4786 4798 4879 4899 4977 5074 5112 5130 5155 5201 5209-5210 5307 5520 5564 5596 5699 5725 5768 5944 5952 5989 6025 6104 6223-6226 6229 6268 6361 6390 6423 6428 6847 6949-6952 7037 7116 7228 7515 7552- 7553 7555-7559 7561 7564 7637 7646 7821 7852 7924 7930 8095 8125 8145 8170 8180 8200 8284 8286 8358 8433 8488 8513 8517 8597 8710 8734 8736-8741 8743-8746 8880 8902-8903 8935 8964 9079 9108 9146 9195 9326 9333 9373 9386 9529 9572 9586 9637 9694 9807 9854 9865-9871 10002 10037 10073 10373 10407 10550-10551 10661 10763 10865 10874 10985 11077-11079 11125 11145 11162 11225 11240 11336 11387 11501 11923 11970 11978-11984 11986-11989 12126 12147 12241 12274 12410 12480 12492 12523 12543 12557 12586 12659 12690 12713 12739-12740 12743 12811 12814 12840 12867 12877 12959 13048 13052 13085-13087 13141 13204 13213 13223 13282 13291 13294 13305-13307 13399 13534 13600 13624 13635 13637 13640 13642 13648 13650 13652 13655 13715 13752-13753 13785 13796 13802 13805 13816 13836 13842 13886 13949 13962 14029 14034 14037 14039 14052
adult brain	GIBCO	ABD003	72 88 147 449 593 604 614-615 617 619-630 639 737 740 743-748 766 806-807 816 1155 1159 1174-1192 1201 1354 1441 1524 1560 1583 1679 1767 1785 1844 1988 2153 2160 2191-2193 2252 2281 2315-2317 2319-2322 2360 2380 2429 2461 2466 2528 2533 2572 2586 2593 2631 2734 2828 2895 2930 2945 3022 3115 3136 3141 3143-3146 3148 3150- 3152 3256 3276-3278 3280-3285 3288-3290 3293-3297 3299-3300 3303-3304 3307 3309 3371 3379 3401 3458 3482 3507 3560 3571 3618 3628 3648 3650 3657 3709-3720 3723 3725 3732 3734 3743 3745 3760 3780 3820 3827 3833 3862 3879 4030 4043 4049 4064 4079 4081-4082 4084 4102 4218 4225 4270 4277 4301 4307 4419 4434 4441 4461 4650 4652 4671 4685 4720 4782-4795 4822 4841 4871 4879 4881-4883 4886-4892 4894-4896 4898-4903 4939 4977 4994 4999 5009 5035 5058 5068 5074-5075 5077 5113 5130 5145

TABLE 12

			12557 12559-12560 12734 12742 12862-12863 12954-12955 13033 13038-13039 13052 13059 13102 13105 13110 13114 13127 13168-13169 13448-13450 13515-13516 13640 13832 13879 13889 13942-13948 13970 14054
uterus	Clontech	UTR001	1043 1141 1254-1258 1332 1501 1503-1519 1630 1679 1940 2160 2281 2326 2343 2360 2384 2409-2414 2572 3304 3502 3743 3780- 3783 3785 3993-4010 4112-4114 4216 4225 4270 4301 4419-4420 4464 4545 5035 5039 5222 5339-5340 5520 5532-5545 5699 5739 5756 5840 5987 6169 6529 6633-6641 6649 6925 7060 7101-7102 7371 7544 7955-7959 8106-8115 8117-8119 8121 8170 8206-8207 8212 8286 8322 8402 8684 8739 9075 9159 9204 9321-9326 9403 9439 9483 9535 9608 9774 10098-10099 10203-10209 10425 10679 10744-10752 10790-10792 11240 11336 11387 11475 11477-11484 11797 11936 12484-12486 12488-12491 12561-12562 12690 12796 12820 12956-12958 13029 13049 13061 13141 13186 13199 13213 13223-13224 13271 13402 13419 13451-13453 13517-13519 13537 13635 13655 13663 13879 13949 13951-13954 14034 14037

TABLE 13

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	M36501	Homo sapiens alpha-2-macroglobulin	222	55
2	A00279	synthetic construct Human serum albumin	332	64.423
3	V00662	Homo sapiens ATPase 6	319	64.583
4	X01683	Homo sapiens alpha 1-antitrypsin	714	80.282
5	M13100	Rattus norvegicus unknown protein	213	36.196
6	M22332	Homo sapiens unknown protein	163	42.391
7	AF015539	Mytilus edulis precollagen P	169	30.682
8	X03325	Homo sapiens apolipoprotein B fragment	742	80.392
9	AB019280	Mus musculus sprouty-4	171	30.872
10	D88152	Homo sapiens acetyl-coenzyme A transporter	818	85.43
11	X92485	Plasmodium vivax pval	167	35.443
12	AB007914	Homo sapiens KIAA0445 protein	289	80.597
13	L35013	Homo sapiens spliceosomal protein	166	33.01
14	L27428	Homo sapiens reverse transcriptase	374	55.085
15	U49973	Homo sapiens ORF1; MER37; putative transposase similar to pogo element	638	74.15
16	U09116	Homo sapiens ORF2, encodes a reverse transcriptase homolog	248	40.625
17	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	162	37.778
18	X53581	Rattus norvegicus ORF4	241	36.364
19	AF183961	Homo sapiens carbon catabolite repression 4 protein homolog	737	80
20	AJ002190	Homo sapiens dihydroxyacetone phosphate acyltransferase	767	85.417
21	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	225	48.227
22	Z74026	Caenorhabditis elegans cDNA EST yk452h4.3 comes from this gene; cDNA EST yk452h4.5 comes from this gene	211	32.331
23	U49973	Homo sapiens ORF1; MER37; putative transposase similar to pogo element	521	66.897
24	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	198	72.34
25	AB013452	Homo sapiens ATPaseII	321	50.394
26	D38595	Homo sapiens inter-alpha-trypsin inhibitor family heavy chain-related protein (IHRP)	782	86.395
27	U49974	Homo sapiens mariner transposase	389	61.475
28	Z92803	Unknown predicted using Genefinder; similar to GTP-binding protein; cDNA EST EMBL:T00705 comes from	177	37.705
29	L07924	Mus musculus guanine nucleotide dissociation stimulator	167	34.959
30	U79260	Homo sapiens unknown	150	61.702
31	U76604	Homo sapiens 180 kDa bullous pemphigoid antigen 2/type XVII collagen	279	30.602
32	AB012223	Canis familiaris ORF2	287	41.045
33	X03145	Homo sapiens pot. ORF VI	606	71.324
34	AC006233	Arabidopsis thaliana hypothetical protein	140	34.959
35	AF003535	Homo sapiens ORF2-like protein	233	50
36	M15386	Homo sapiens gamma-globin	672	83.704

TABLE 13

		comes from this gene; c		
14067	AF107406	Homo sapiens GW128	122	69.565
14068	AF078851	Homo sapiens secretogranin III	2980	99.145
14069	U62039	Elephantulus edwardii reverse transcriptase	139	48.718
14070	AB007861	Homo sapiens KIAA0401	2399	100
14071	X07816	Human herpesvirus 4 epitope Cl3 (57 AA)	69	53.846
14072	S58722	Homo sapiens X-linked retinopathy protein {C-terminal, clone XEH.8c}	128	45.455
14073	AB006136	Homo sapiens FUT1	136	86.364
14074	L19684	Homo sapiens kallistatin	1128	43.972

TABLE 14

SEQ ID NO.	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of amino acid sequence	Predicted end nucleotide location corresponding to last amino acid residue of amino acid sequence	AMINO ACID SEQUENCE ENCODED BY THE NUCLEOTIDE SEQUENCE OF THE CORRESPONDING SEQ ID NO. (A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=possible nucleotide deletion, \=possible nucleotide insertion)
1	A	2	199	LKGVNSPIKKPRVASFFKKNPRPLVFCF*K/T/HLFCKPPGVQKK GWGKIFPANWKP KKTGGGLNFDKTA*PPKF
2	A	1	2321	FFKNFFPPRVWGGFFSPFSP*KSSSPKAFIFLGGVGPIFPPPKKRF FFKNSQGGVFPF/PPVFKNRPRPCFFLPPPPSSSSSSP\PPVNFPGP PRVFFKGPPSSSSSSSSSLRQS
3	A	2	517	YAKLGTRILRLVKNPPAATGGCAPPASEAQASSQFALSSALYLPGE \QSMVPVEKMRSAGSRG*WRRAA*QPWCSGRAPPWPWAASATTS SSATA\LPQLSDGH*T/SVPPTPPHSPPLGDCSQRPGNWQSEA* APGQLPGTSLVPLGGSQPSQPCPLPPLSPFLFPFPA
4	A	471	911	KWAGTGAGAPDPLQSGLVTTPTQPGFRPTL/PAPPCSGLP CPRAPP WYTPSQAGDPPPRTPQAADAQEHRARPCPPSAGVLGPVPTCFCPQP ALSP*LHPWPT*KVP SHALQPAKALAHLT LHGQCHHASHVP*AGS HCSCCEFPDT
5	A	1	1250	ARTFPLALDEAAGRGAAEGPAALLG\PAGFRWAEPGAALGRGLAA VVG RGATWRRRGRCPAGRIVPSVPARCALLPPSGAAGQGQRLRQ RRR*AGRALQPGSGQLRPCPA/PPGVRRPQPRGAPGQS AFCSPPA LDKLRLCGSATPKARPGEAAQRVGSDLGSSGPRGRLSLPSSVYPS WSVP\PGPSGVPGVPGGADCS/P*QEGRGQGTDDPEVPALNEQAAP TSWQASFWDRGDIAGKCSGSRGLRKESASSGLDISTPQHSSG*SD P*LAPGHL*GSQAAGDRWPGRPLLPAGTVAP/GSGNQSPPTMCGA PGD*VANGKPPCFSGARAAS\GSEETPLPSAPHLSLGDTRAPYPGQ *WRGISTGVLGLHLLPSLHVFPAPPSARARPPGRSFPASGLCPPAS RRPGNSGQ
6	A	31	319	SIVWTATLFLLEQKGT LKMTDQTSTSQMMTQKVNLLKRAKERQIS* GR*RECK\IQRPCRKS VNKMLFVLVLEFAICWAPFHIDRLFFSFVE EWSE
7	A	125	421	GGEPYRRNEDKPVA\ACGAIANSVFNDTLKKVLIGFDW*PIPMGWK KNGIAWRTDKRVKFKNPPGGDNLGERFKGTSVPGNWLKPAYMLDSE PNNNGFI
8	A	2	319	FFLRWSL/DSVAPAKVQWRDLGSGQAPGFTPFSCLSLPSSWDYRHP PPRPANFFVFSVEMGF\TMLARMISIS*H\VDPPTSVSQNAGITGM SHRASLDILLNVLLS
9	A	1	729	LAVKMALSRVCWARTAVLGSAVTPGHFVTRRLQLGRSGLAWGAPRS SKLHLSPKADVKNLMSYVVTKTKAINGKYHRFLGVVIFPRFLY\LY TIFMKGRIADVGM\DAKKG*RNKAQLCEGQL*GFINFQSREMEHLR QFRQDVPKCLFIGIISIPPFANYLVFLLMYLFPRQLLRHFWTPKQ QTDFLDIYHAFRKQSHPEIISYLEKVIPLISDAGLRWRLTDLCTKI QRGTHPACMRPLR
10	A	1	283	KSVYCW*\NIQLIKPFWKAVWHSVLKLNIFIFCNLVFPLLDI*SKL CSCVPGARIKIFTALLFILTLIT/VNKNLEHL*HPAEVEWVNQFWY IYIY

TABLE 14

[illegible]

TABLE 15

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	2 4 17 77 140 433 508 828 952 1084 1125 1134 1183 1446 1555 1615 1689 1788 1803 1865 1886 1931 1945 2031 2058 2154 2453 2689 2724 2738-2739 2772 2787 2920 2941 2967 3106 3160 3177 3179 3227 3242 3321 3356 3360 3369 3402 3412 3424 3432 3446 3449 3453 3492 3658 3735 3749 3756 3789 3791 3889 3977 3982 3986 3989 4006 4180 4195 4198 4205 4251 4255 4290 4357 4416 4427 4429-4430 4556 4596 4600 4602 4611 4672-4673 4676 4722 4752 4759 4777 4828 4860 4885 4906 4939 5096 5110 5117 5130 5154 5321 5326 5342 5350-5351 5362 5365 5371 5376 5379 5385 5389 5392 5403 5425 5427 5457-5458 5467 5482 5501 5505-5506 5508 5511 5514-5515 5522 5531 5539 5543 5548 5557 5573 5600-5602 5606 5615-5616 5622 5632-5633 5638 5640 5646 5648-5650 5654 5657-5658 5663-5664 5668 5674 5679 5681-5682 5686 5691 5695 5698 5703 5710-5711 5713 5718 5720 5724 5727 5730 5732 5734 5738 5749 5751 5763 5767 5782 5789 5795 5824 5833 5869 5878 5881 5889 5893 5899 5905 5910 5913 5931 5946 5984 6022 6041 6059 6141-6142 6146-6147 6154 6222 6255 6292 6298 6313 6346 6364
adult brain	GIBCO	ABD003	2 4 9 17 77 130 180 268 297 371 433 508 547 643 657 689 720 782 785 801 809 811-812 828 941 947 952 1023 1127 1139 1145 1154 1157 1160 1364 1421 1438 1544 1592 1597 1623 1628 1646 1650 1659 1661 1663-1664 1692 1695 1779 1793 1803-1804 1848 1852 1925 1980 2035 2059 2080 2121 2453 2469 2496 2512 2534 2555 2581 2595 2621 2637 2689 2720 2724 2733 2737 2746-2747 2753 2763 2767 2774 2796 2845 2901 2945 2954 2967 2992 3020 3064 3077 3091 3102-3103 3106 3133 3177 3179 3183 3187 3191 3204 3206 3232 3240 3242 3246 3252 3295 3316 3331 3351 3368 3376 3378 3401-3402 3405 3409 3412 3432 3434 3436 3446-3447 3453 3456 3460-3461 3484 3492 3527 3636 3639 3653 3658-3660 3662 3724 3735 3746 3748 3758 3768 3771-3772 3789 3791-3792 3808 3867 3889 3913 3950 3965 3986 3989 3991 4006 4008 4010 4014 4017 4041 4043 4048 4082 4112 4167 4188 4195 4198 4205 4212 4219 4226 4228 4230-4231 4238 4241 4243-4244 4246 4252 4255 4263 4266 4270 4275-4276 4282 4291 4293 4297 4347 4416 4427 4429 4445 4459 4462 4493 4541 4544 4546 4558-4559 4563-4564 4581 4591 4596 4600-4602 4608-4609 4611 4617-4618 4639 4655 4665 4673 4676 4722 4741 4744 4749 4761 4764 4828 4860 4864 4906-4909 4960 4962-4963 4981 5017 5038 5047 5067 5098 5117 5123 5127 5147 5151 5223 5250 5256 5259 5263 5273 5276 5288 5320-5321 5325 5327 5330-5331 5336-5337 5341-5342 5347

TABLE 15

			3889	3982	4112	4182	4238	4244	4297	4376	4421
			4459	4505-4506	4559	4596	4777	4891	4906	4963	
			5076	5110	5123	5273	5313	5350	5368	5387	5403
			5448	5463	5477	5498	5513	5523	5531	5569	
			5609-5610	5613	5622	5630	5633	5652	5675	5683	
			5695	5701	5704-5705	5710	5717	5720	5730	5737	
			5747-5748	5758	5761-5763	5767	5782	5789	5798		
			5841	5854	5878	5884	5891	5903	5929	5935	5938
			5944	5946-5947	5949	6041	6118	6152	6183	6216	
			6241	6255	6285	6302	6375	6380			

TABLE 16

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	U93564	Homo sapiens putative p150	156	44
2	M76546	Helianthus annuus hydroxyproline-rich protein	169	36.364
3	Z98980	Schizosaccharomyces pombe wiskott-aldrich syndrome protein homolog 1	163	34.94
4	L21990	Homo sapiens spiceosomal protein	198	32.576
5	AF153062	Canis familiaris type I collagen pre-pro-alpha1(I) chain	270	31.264
6	AF044601	Homo sapiens orphan G protein-coupled receptor; GPC-R	180	36.17
7	U61953	Caenorhabditis elegans No definition line found	175	44.595
8	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	322	67.708
9	Y10018	Drosophila melanogaster ANON-66Db	185	25.568
10	M13100	Rattus norvegicus unknown protein	140	33.333
11	U94492	Meloidogyne javanica collagen	163	34.286
12	AF055985	Onchocerca volvulus pyrrolidone-rich antigen	141	41.667
13	M80344	Homo sapiens ORF1 codes for a 40 kDa product	184	35.227
14	D10493	Homo sapiens p64 myc protein	296	43.151
15	D10493	Homo sapiens p64 myc protein	318	48.062
16	M22334	Homo sapiens unknown protein	191	55.294
17	U67056	Acanthamoeba castellanii myosin I heavy chain kinase	191	34.872
18	M80341	Homo sapiens ORF2 contains a reverse transcriptase domain.; ORF2	242	28.986
19	AF003535	Homo sapiens ORF2-like protein	243	51.685
20	D80009	Homo sapiens KIAA0187	815	92.537
21	D88460	Homo sapiens N-WASP	877	88.636
22	L02918	Mus musculus procollagen type V alpha 2	144	35.503
23	L24521	Homo sapiens transformation-related protein	288	51.754
24	AF003535	Homo sapiens ORF2-like protein	150	41.096
25	AF003535	Homo sapiens ORF2-like protein	243	41.88
26	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	157	42.647
27	AF003535	Homo sapiens ORF2-like protein	175	49.206
28	AF003535	Homo sapiens ORF2-like protein	144	49.315
29	AB012223	Canis familiaris ORF2	207	56.164
30	AF118023	Homo sapiens SH3 domain-binding protein SNP70	182	38.462
31	X77816	Rattus norvegicus PR-Vbeta1	152	30.508
32	X83413	Human herpesvirus 6 U88	255	52.688
33	X83413	Human herpesvirus 6 U88	221	52.055
34	M88593	Gallus gallus alpha-1 type XI collagen	187	28.526
35	X05561	Homo sapiens alpha-1 chain precursor (AA -27 to 917) (2953 is 2nd base in	170	40.541

TABLE 16

6369	AB011540	Homo sapiens MEGF7	10912	100
6370	U22456	Homo sapiens AMP-activated protein kinase homolog	175	34.884
6371	AE000762	Aquifex aeolicus hypothetical protein	165	24
6372	AF159055	Homo sapiens leucine zipper-like protein	123	64.516
6373	D30747	Acropora donei mini-collagen	127	33.333
6374	X52164	Mus musculus Q300 protein (AA 1-77)	131	38
6375	AB014516	Homo sapiens KIAA0616 protein	4273	100
6376	A31036	Nicotiana glauca PRP2	130	31.579
6377	M12623	Homo sapiens high mobility group protein 17	351	93.103
6378	D29833	Homo sapiens proline rich peptide P-B	75	36.111
6379	Z70292	Homo sapiens chemokine CC-1	207	100
6380	M94131	Homo sapiens mucin	9126	100
6381	M15885	Homo sapiens seminal plasma protein precursor	801	99.123
6382	M22865	Homo sapiens cytochrome b5	603	76.119
6383	AF010427	Hepatitis E virus ORF-1; hypervariable region	118	58.621
6384	AJ131190	Homo sapiens FANCA protein	134	66.667
6385	S79410	Mus sp. nuclear localization signals (NLS)-binding protein=spot-1	160	57.895
6386	AJ388513	Canis familiaris Ribosomal protein L6	113	85
6387	AF039052	Caenorhabditis elegans No definition line found	183	29.299
6388	X67703	Drosophila melanogaster Mst84Dc	93	48.148
6389	AJ388550	Canis familiaris hypothetical protein	112	41.667
6390	X01779	Hordeum vulgare C-hordein fragment	128	33.333
6391	AL109822	Schizosaccharomyces pombe hypothetical protein	117	26.027

TABLE 17

SEQ ID NO	Method	Predicted start nucleotide location for amino acid sequence	Predicted stop nucleotide location for amino acid sequence	Amino acid sequence encoded by the nucleotide sequence of the corresponding SEQ ID NO. (A=Alanine, C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=Possible nucleotide deletion, \=Possible nucleotide insertion)
1	C	23	76	MNNLILRRHLNKNYFCY
2	B	1	735	MSGDISGNMFLREEGIQNTRLYTAKPVDLVMAMDLHQSVARRLGTP GVTSCPSYNEKLVDPDKLKIRYHSSDFQKSILCNTLTIMDQLVKGGAI MVKGLTQPCSGLSVCQNQHTIRTSGFFRADPKISYIPDVLMSNRQSS EDMLLIPTRRNLPRIQAKNND SQVPMKSWMDVELGNVLADSRGRCE PQSSRAGTSRERRSLPPIYGLQLLCPVIFGWPLPTGSTGRKSEDEKR DSK
3	B	1	783	MPSEERRAGLPPSLPDRGRAGEHDGSSNSRGGAGRQARSARRHSEAI RAAGSAQKYDQRD LHANIPNSIVHANQNVHQPRMHEQQGNGHEKKLL RHGPTWMNLKNARPAASIELHHLFGVHNRTQ RATRLQTPLPSRAPRT NIRQGAPYLVGSTLSHLGSGKRPEGGRGSSGQSGVGSRAERIDGA PLPRPPLPAFLNAFVFLDPGTQPRAPTGGNSSQGLLSPARGFTGAPGP GTSFSATTTCQETPPQLPICPQMF
4	B	104	266	XIPYHTPWNP RQEKYDQRD LHANIPNSIVHANQNVHQPRMHEQQGNC PWKGTAETWAYMDEPQKR
5	B	1	1113	MKRPNSDLSTNCSGTVCLDPAYGEEDEGLPGRLPGEGLREPTCD GWTS LQAAGWMSVHTGGSQEMRLQGQKIAMANAEC LCRARPCASAL SCPTAQQA RHPDTILPRQLGHWKAVTRVKLLNSYHGRSPFTTLRHST RKESIWNKQSIANQNALFQFPTQVSLTAFITGLKRPCKSFSASTGCRFLI VTTGMVSAVQPETQDQGLRSFDTIWQLRAQRNSIISVNRINSLOGLCI YYQKGINTDLKSTYFQKPDAGRSHSGSRSHKLRT EPASSQHKGRC SK TSSIGGLAVELMLLTVSV PWSQSMSQATQGPASPCSPFVKRAEPPE CLLGKWL PSCRLISSAPPQLGVPTKSLFPQADKQTK
6	C	3	164	MFRIQYFWYVPPPTNRHMMHMIHFQQRYS EDHTSRTISKASVTLSLSR SPGEQ
7	B	112	279	MSPYKGVAGPASRCGECGGARKLRDASKHGAPREQKHCQSLLWSG RQECVTACSL
8	B	198	405	EKEEEENVEEILRGDWP NRQVSMGCWDTLAVRETET
9	B	1	687	MSHRAWPPCCTGPNAPSQVPLVLEDVSNVYGDVEIDRNKHIHKKRKL AEGREKTMSSDDEDCSAKGRNRHIVNKAELANSTEVLESFKLARES WELLYSLEFLDKEFTRICLAWKTDTWLWLRIFLTDMIYQGQYKAIASL HHLAALQGSISQPQITGQGTLEHQRALIQ LATCHFALGEYRRREGMGIK TGAVAVDLLKVVKDMIQGIGNLTTLFHFPLLGD F
10	C	346	600	MLLSCSSW/LVASCWQTLSDIFSGGKTKCISLGCQINSALTSFHTALEL AVDQREIQHVCLYEIGWCSMIELNLKDAFDSFERL

TABLE 17

30468	B	1	714	MAFGLGAIWGVILITCLLPVNQLLTALPVDVLGSLGELSSPVVSFAFALFP LVAIFYQFGWKQSLIAAVVLMTRVVVVRYFPHLNPESEIEFIGMVMLLGI AITHDLRHRDENDIDASGLSVFEERTSRIKNLPYIAIVGALIAAVASMKIF AGSEVSIFTLEKAYSAGVTPEQSQT LINQAALAEFMRGLGFVPLIATTAL ATGVYAVAGFTFVYAVDYLSPNPMVA AVLGA VVI
30469	B	166	1337	XILSRMYGNLPIEYGADGAGSGNCGTPCRVTP TERRMEVYSISIWMMT PCVFHMMGSARNASYCDGITSRRLYLIWGMVGENQV FDDMDQWPLKI CASCCHKRIRYVMILSAFSLEGKVA VVMVILDWKVNVDFVMLTEYAA DSKLRLLTHPLPWLT PSGVDGEKRQNSVAYFRNRKYAHKSRQRRRSH SDQLTESGYLDLQPEMSSAELTEKLVGPSRAANTTPSSVTIRTRHDDQ VLTQLIVVWKL PARIWSLKKLHYSRPNSSGGATRHRKTILSNTGDDLDVI GVAFDAAWLVSTYCSFVRAKCSQPQLFPESAWRYGTERVWETFVQGQF YLQGSQMRPYRNWRDARLMFNQTRGD RARYLKLARTNAATALTSKLS QQSTVHRD
30470	B	72	617	MGPNVSGNPVIKYFNLYTYKMPGDIKNWVDAHMNCEDIAMNFLVAN VTGKAVIKEAGVKKPQDRSSSDLPSPLEKADSESNKGKKRRNRRTFTS YQLEELEKVFQKTHYPDVYAREQLAMRTDLTEARVQAVIRQRAGNARK ETNRAQVDVLVEFTTKFQQRT PQRN MVRNGIRPPTAPK
30471	B	1	1002	MSALITSSTPALGGRSLPLCAGPAGYTA AVYAARANLQPV LITGMEKG GQLTTTTTEVENWPGDPNDLTGPLL MERMHEHATKFETEIIFDHINKKPL KAVGVSACATCDGFFYRNQKSSGHRPAAIPRVKEALYLSNIASEVHLIH RRDGFRAEKILIKRLMDKVENGNIILHTNRTLEEVTGDQMGVTGVRLRD TQNSDNIESLDVAGLFVAIGHSPNTAIFEGQLELETILRNPRPSPKFKNQ LSCFPGKIVTVPFPTNINRIKIKDSLKAAPKSQKADSPSIDY AELLQHFEK VQNKHLEVRHQSRGRGDHLDRRVLTGLARRRAHRVVP
30472	B	1	4173	MELGRGIKGLIGHTQSCRLAARTVANRIAEDLKHGAGRLHRLQSAFQR SPEEDTNVVRHNVPGA AHSRTFRTFSGIAGGWITTLA IKKRQCTRA CIKQTA FMDSMKSDMGGAATVTGALAFAITRGLNKRVKLFLCCADNLIS GNAFKLGDIIYRNGKKVEVMNTDAEGRVLADGLIDASAQKPEMIIDAA TLTGA AKTALGIASMHCSANVGEKGDVAVFFGLSGTGKTLSTDPKRR LIGDDEHG WDDD
30473	B	1	4488	MAVAAAALSLLLLLLILSRTNEFDLRVFIQPQHDSGEDTPYWSQIIVTIII TWDQMAPTSKGYHEQSMISYASPEKAIRQLNQSYNVGKGPDPYPAGKE GLHHIKAENRLSNAVTSATSPPLKKLPLARTANVDVAEYRRDCGSRFR SYAIFTGHFFHFIARHPTPPWRTNWR CNWFCRRSPARKAAALAKASV NKVFFIQPTPQFVIGQTEAGIRIISQLPLRTHIPGYGAAAF AQPGRAIA MTINREPQRG
30474	B	1	3822	MKIPSLVGPDGKTSLREYAGYHGGGGGFGGQLRGWNPPSESADAAL LPNFSRGNARADDLVRNNGYAANAVQLHQDHIVGSFFRLSYRPSWRY LGINEEDSRAFSRDVEAAWNEYAEDDFCGIDAERKRTFTMMIREGVAM HAFNGELCTQATWDS DSTRLFRTQFKMVSPKRVSNNPNI GDTRNCRA GVKINDSGAALGYVSDDGYPGWMAQNWYI PRELPGGRPSFIHVFE PMEDGQTRGANAFYSVMEQM

TABLE 18

adult brain	GIBCO	AB3001	39-41 199 204-207 322-323 549-551 585-589 617-631 644 1015 1199-1201 1287-1293 1452 1498 1746-1747 2066 2167-2168 2489 2503-2506 2545 2631-2635 3022-3023 3140-3141 3249 3313-3314 3343-3345 3389 3402 3411-3413 3470 3485-3487 3592-3595 3733 3990-3994 4108-4113 4201-4203 4599-4601 4622-4623 4653-4655 4878-4881 5529-5530 5649-5652 5746 5848 6026-6030 6184-6185 6453-6454 6502-6503 7263-7266 7501-7503 7506-7514 7590-7595 7608-7609 7676-7677 7684-7690 7714-7719 7755-7757 7825 7832-7837 7852-7855 7976-7978 8049 8065-8067 8080-8081 8200-8206 8282-8284 8472 8554-8556 8995-8997 9045-9050 9061-9063 9072-9076 9397-9398 9426-9428 9435-9436 9451-9452 9455-9456 9491-9496 9503 9535-9536 9577-9578 9582-9585 9646-9649 9663-9664 9912-9913 9915-9916 9935-9956 9966-9969 10008-10013 10030-10041 10047-10053 10086- 10087 10094-10098 10228-10233 10338 10541- 10543 10549-10550 10559-10564 10581-10583 10598-10599 10653-10656 10690-10692 11294- 11295 11407-11412 11474-11475 11499-11500 11802-11805 12224-12225 12313 12315-12318 12631-12635 12691-12694 12863-12865 13058- 13061 13064-13065 13077-13079 13647-13648 13672-13675 13686-13688 13953-13955 14081- 14086 14446-14455 14736-14737 14770-14771 15100-15105 15183 15264-15271 15299-15300 15340-15341 15347 15413-15414 15638-15642 15658-15662 15666-15667 15678-15679 15790- 15791 15972-15974 16535-16536 16570-16575 16593-16594 16709-16711 16787-16794 16801- 16802 16828-16830 16940-16946 17038-17041 17053-17057 17561-17562 17963-17967 18064- 18069 18136-18137 18243-18245 18530-18532 18623-18625 18644-18646 18733-18735 18777- 18781 18828-18831 18859-18865 18899-18902 18911-18913 18945-18947 19008-19012 19028- 19030 19052-19054 19056-19059 19073-19078 19098-19099 19122-19126 19154-19157 19166- 19174 19251-19256 19263-19264 19336 19429- 19430 19458-19462 19468-19473 19475 19483 19498-19502 19528-19530 19535-19544 19555- 19557 19618 19675-19677 19681 19719-19722 19808 19824-19825 19842-19847 19851-19857 19887 19919-19923 20037-20045 20049-20054 20059-20062 20085-20097 20131-20133 20145- 20159 20203-20210 20215-20218 20227-20228 20238-20243 20277-20280 20283-20287 20296- 20297 20305-20310 20314-20315 20331-20334 20397-20399 20406 20438-20441 20466-20471 20478 20510-20517 20532-20534 20554-20557 20641-20652 20659-20662 20671-20675 20725- 20730 20732-20733 20764-20770 20816-20825 20836-20843 20845-20850 20908-20910 20925- 20931 20963-20968 21011-21012 21049 21061- 21065 21125-21130 21153-21157 21189-21193 21199-21202 21264-21267 21272-21275 21280- 21281 21400-21402 21405-21411 21538-21542 21562-21567 21613-21619 21623-21628 21775- 21783 21857-21861 22057-22063 22086-22089 22101-22102 22106 22128-22134 22154-22157
-------------	-------	--------	--

TABLE 18

			17560 18064-18068 18122-18123 18635 18764- 18767 18782-18786 18870 18898 18934-18943 19003-19005 19008-19012 19045-19048 19145- 19148 19183-19192 19475 19483 19488-19492 19500-19502 19555-19557 19675-19677 19724- 19730 19808 19842-19847 19879-19882 19932- 19934 20042-20045 20049-20053 20066 20097 20145-20159 20236 20238-20243 20262-20264 20267-20270 20406-20413 20415 20445-20447 20483-20485 20518-20522 20544-20548 20586- 20588 20608-20611 20671-20674 20720-20724 20747-20748 20767-20770 20807-20810 20871 20876-20886 20920-20924 20977-20981 20983 21061-21065 21131-21134 21199-21202 21340- 21342 21409-21422 21505-21525 22039-22040 22083-22085 22106-22110 22147-22153 22178- 22183 22218-22219 22315-22320 22346-22352 22379-22380 22389-22392 22486-22487 22490- 22492 22501 22533-22536 22699-22709 22750- 22752 22772-22779 22791-22792 23019-23023 23088-23090 23102-23103 23181-23184 23228- 23231 23275 23560-23562 23574-23582 23678- 23679 24167-24171 24268-24282 25223-25228 25480-25481 25515-25516 25557-25558 26368- 26370 26417-26419 26432-26437 26474 26758- 26760 26828-26829 26908-26909 27006-27007 27013-27015 27336-27337 27457 27511-27516 27658-27661 27768-27772 27805-27808 27818- 27823 27898-27908 28031-28034 28066-28097 28275-28291 28303-28307 28483-28485 28598 28600-28602 29514-29515 29555-29556 30145- 30146 30334-30340
--	--	--	---

TABLE 19

SEQ ID NO	ACCESS-ION NUMBER	DESCRIPTION	SMITH-WATER-MAN SCORE	% IDENTITY
3379	AJ133271	Gorilla gorilla atrophin-1	51	29.73
3380	AJ133271	Gorilla gorilla atrophin-1	149	27.536
3381	AJ133271	Gorilla gorilla atrophin-1	149	27.536
3383	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	344	60.825
3384	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	634	87.85
3385	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	369	82.353
3386	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	523	38.776
3388	Y12713	Mus musculus Pro-Pol-dUTPase polyprotein	467	68.807
3390	AL049758	Homo sapiens dJ437M21.1 (novel protein)	237	70
3391	AL049758	Homo sapiens dJ437M21.1 (novel protein)	344	47.682
3392	Y09443	Homo sapiens alkyl-dihydroxyacetonephosphate synthase precursor	1266	85.965
3393	Y09443	Homo sapiens alkyl-dihydroxyacetonephosphate synthase precursor	1243	92.991
3394	X79536	Homo sapiens hnRNPcore protein A1	1959	91.875
3395	X79536	Homo sapiens hnRNPcore protein A1	593	51.261
3396	X79536	Homo sapiens hnRNPcore protein A1	315	95.652
3397	X79536	Homo sapiens hnRNPcore protein A1	1486	98.4
3398	U34360	Homo sapiens LAF-4	8177	100
3399	U34360	Homo sapiens LAF-4	205	59.322
3400	AL033534	Schizosaccharomyces pombe serine-rich protein	165	30.46
3401	AL117470	Homo sapiens hypothetical protein	281	40.764
3402	U22376	Homo sapiens alternatively spliced product using exon 13A	313	69.231
3403	L11672	Homo sapiens zinc finger protein	2454	70.623
3404	AB006628	Homo sapiens KIAA0290	288	46.939
3406	AB006628	Homo sapiens KIAA0290	183	36.029
3407	AL110124	Homo sapiens hypothetical protein	604	64.216
3408	AL110124	Homo sapiens hypothetical protein	219	53.684
3409	AL110124	Homo sapiens hypothetical protein	204	94.595
3410	AL110124	Homo sapiens hypothetical protein	3778	97.879
3411	X56932	Homo sapiens 23 kD highly basic protein	820	80.663
3412	X56932	Homo sapiens 23 kD highly basic protein	505	100
3413	X56932	Homo sapiens 23 kD highly basic protein	1260	98.515
3414	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	289	39.726
3415	AL132896	Caenorhabditis elegans predicted using Genefinder; preliminary prediction	257	26.621
3416	Z72499	Homo sapiens herpesvirus associated ubiquitin-specific protease (HAUSP)	867	92.308
3417	Z72499	Homo sapiens herpesvirus associated ubiquitin-specific protease (HAUSP)	557	32.718
3418	Z72499	Homo sapiens herpesvirus associated ubiquitin-specific protease (HAUSP)	7326	99.637

TABLE 19

30403	AF039023	Homo sapiens Ran-GTP binding protein; RanBP6	1020	86.188
30404	AF039023	Homo sapiens Ran-GTP binding protein; RanBP6	7063	97.922
30405	AC002400	Homo sapiens Gene product with similarity to Ubiquitin binding enzyme	1536	95.102
30406	AC002400	Homo sapiens Gene product with similarity to Ubiquitin binding enzyme	341	90.741
30407	AC002400	Homo sapiens Gene product with similarity to Ubiquitin binding enzyme	3437	98.876
30408	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	344	60.63
30409	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	3687	100
30410	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	3633	98.611
30411	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	46	40.909
30412	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	1311	85.156
30413	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	1369	95.671
30414	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	2808	79.701
30415	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	3357	96.701
30416	M24903	Homo sapiens gamma-glutamyltransferase 1 (EC 2.3.2.2)	2944	82.89
30417	U02390	Homo sapiens CAP2	80	36.364
30418	U02390	Homo sapiens CAP2	346	59.664
30419	U02390	Homo sapiens CAP2	2204	91.02
30424	AF068864	Homo sapiens p21-activated kinase 3	137	55.263
30425	AF068864	Homo sapiens p21-activated kinase 3	260	78.182
30426	AF068864	Homo sapiens p21-activated kinase 3	2883	87.184
30427	AF121080	Mus musculus cAMP inducible 1 protein	518	85.106
30428	AF121080	Mus musculus cAMP inducible 1 protein	376	79.452
30429	AF121080	Mus musculus cAMP inducible 1 protein	332	87.5
30430	AF121080	Mus musculus cAMP inducible 1 protein	59	43.75
30431	AF121080	Mus musculus cAMP inducible 1 protein	683	58.019
30432	AF121080	Mus musculus cAMP inducible 1 protein	54	25
30433	AF121080	Mus musculus cAMP inducible 1 protein	60	28.986
30434	AF121080	Mus musculus cAMP inducible 1 protein	1323	53.012
30435	AF121080	Mus musculus cAMP inducible 1 protein	2833	72.205
30436	U17195	Homo sapiens A-kinase anchor protein	5612	99.424
30437	U17195	Homo sapiens A-kinase anchor protein	3841	99.831
30547	X98743	Homo sapiens RNA helicase	728	62.185
30548	X98743	Homo sapiens RNA helicase	2996	94.059
30550	X98743	Homo sapiens RNA helicase	689	90.517
30551	X98743	Homo sapiens RNA helicase	1275	75.986
30552	X98743	Homo sapiens RNA helicase	57	21.25
30553	X98743	Homo sapiens RNA helicase	3082	88.694

TABLE 20

SEQ ID NO.	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of amino acid sequence	Predicted end nucleotide location corresponding to last amino acid residue of amino acid sequence	AMINO ACID SEQUENCE ENCODED BY THE NUCLEOTIDE SEQUENCE OF THE CORRESPONDING SEQ ID NO. (A=Alanine, C=Cysteine, D=Aspartic Acid, E= Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=possible nucleotide deletion, \=possible nucleotide insertion)
1	C	239	322	MGGALLKEPILSPGGGKGKIFFWGPQN*
2	A	1441	2130	SVIA*SCRASVASKQS*PTLLPSACARPHA\STVDAPASGGAPRASSP\SSD CLWSTSSSSTPLSASASSS/SPPSFNPAADARGSQGPARGRSCSPSSSERH VRRRVSAARQAGAASAGGGRQAGLAGRSGLSA/SRSSARASSSATPALAQST \PSESECAPLKSRSGLTSSLSKPAS*ATLGKKGSGSSWRFPPESIHGRHPL SASCWNKSVAAAAAPTATAPPKAGP
3	C	36	236	MGPTIPDXSFFWRKPITWMPTWEGTSNVGPQPLSSSKSLHSXRGHPAPIPT GQAGPRDSGPGASP*
4	A	109	300	GGGKQIPFKGGKFKWGPVLLKKGEREKPGGNPKKTPWKKASSRPAPRIHPC FT*HAPDPRPLY
5	A	2	74	FPGKFFFGYPRFFFPFLNPGPG
6	A	27	375	EHSGVRQALCFGTASQRPSQQPAPSGPGPPGEPG*ERLCASHKAFISHKQSH *SPQ*PCQAGVTLSRLQTTSNRPHSQKGLRGPRQTLSLTSTQPTACSENSQ GSQPSPKRTLS
7	B	50	204	XKEGSLCDEYWNPAANLINVCSLFLRQGPRLALMQGEPVDKGCGLGVLENK*
8	A	379	623	ATTVSVPFPTAKLLERPGHLHLLVFLPNLQFPLQPLVS*LALLRGSTLTQVVP SAPDKPLLVSPPAKHPPVPPSCGPGQLQG
9	B	185	366	XHPGDGFRPNQEGDERPARKKTWVRDGGPHQGLFRSFHPQFFSRPSRATAHV PAVYFSVEWX*
10	A	29	308	WLPPNPGRRRREARQEEDLPGWGWAPSGPLPQLPSAVLQPTQPGHGPRASL** SVCFSFADKEGSLCDEYWNPA/KPH*RLQPLPSTRPEISPL
11	A	663	1270	TAGTWAVASLGRLLKNCGWKLKKEALMGPTIPDPKSSPLAGLSSFPWFGRKP ITLCEPTWERDPRNVGPAPSPARKSLPQPTGTTLQPYSPRDKAGPKKTLGP RG/APL*VRRTRPLN*WTPADLGVTRTRGAGPLPDAGTLRPRGAVEPSVSAC GKWAPSPTSQGCCEGRCDVAPKHEGLAHTVLSINVFPVLNQKKKKK
12	A	190	715	WRWRWRRKPTARTASTRHIFLHIKEALLAGHLLCSPEQAVELSALLAQTKF GDYNQNTAKYNYEELCAKELSSATLNSIVAKHKELEGTSQASAQVQLQIVS AMENYGIWHSVRDSEGHKLLIGVGPEGISICRDYFSPINRIAYPVVQMATQ SGKNVYLTVTMESVNSIEL
13	A	270	713	KLTIILDCQFTG*QR*KFNG*NLRNR/HSPSRWDGAKPLYKALKL*SSSSSVG AFIFIFITRSRLRAYLFSFAH/LRRPLLAGHLLCSPEQAVELSALLAQTKFGD YNQNTAKYNYEELCAKELSSATLNSIVAKHKELEGTSQASAQVQL
14	A	1575	1968	REMGRFHVGGTGLELLTSGDLPTSASQSAGITGVSHHTWPKTLFVLRQSLTL SPGLECSGTISAHCSPHLPCSSNSCAPASRVAESTEAAH/LCPDNLHISRE GASPCWPGCS*TPELKRAHPCRDQLGH
15	A	185	721	KVSHVYFLHRHGNHPISTQFPHLSPLSIPQNCCHHGPFMSWCWRIKYLGI QLTRDVKDLFKEN*KPLLSKIKEDTNKWNIPCSWIGRINIVKMAILP/KEL EKTTLKFIWNQKRACIAKTILSQKNKAGDITLPDFKLYYKATVTKTAWELQ RDIDQWNRIEPIPHIYNYLIF

TABLE 20

				MHINPANKSLPPTFGPATLFNHFS S LFDSSQVPANQGWGDGPLSSRVATDAS FTVQSAFLGNSVLGHLENMHPDNSKAPGFRPPSQRVSTSPVGLPSIDPSGSS PSSSSAPLASFSGIPGTRVFLQGPAPVGTSPFNHQHFSHPWTSASNSSTSA PPTLGQPKGVSASQDRKIPPPIGTERLARIRQGGSSVAQAPAGTSFVAPVGH GIWSFGVNAVSEGLSGWSQSVMGNHMPHQQLSDPSTFSQHQPIMERDDSGMVA PSNIFHQPMASGFVDFSKGLPISMYGGTIIPSHPOLADVPGGPLFNGLHNP PAWNPMIKVIQNSTECTDAQQIWPGTWAPHIGNMHLKYVN
5818	A	2	1815	GGTESTRYAVQLINALIQDPAKELEDLIPKNHIRTFASTKSIHANFSSSGVGT TAASSKNAFPLGAPTLVTSQATTLSTFQPANKLNKNVPTNVRSSFPVSLPLA YPHPHFALLAAQTMQQIRHPRLPMAQFGGTFSPPNTWGPFPVRPVNPNGNTN SSPKHNNTSRLPNQNGTVLPSESAGLATASCPITVSSVVAASQQLCVTNTRT PSSVRKQLFACVPKTSPPATVISSVTSTCSSLPSVSSAPITSGQAPTTLFLPA STSQAQLSSQKMFSAVPPTKEKVSTQDQPMANLCTPSSSTANSCSSASNT PGAPETHPSSSPTPTSSNTQEEAQPSVSDLSPMSMPFASNSEPAPLTLTSP RMVAADNQDTSNLPQLAVPAPRVSHRMQPRGSFYSMVFNATIHQDPQSIFVT NPVTLTPPQGPAAVQLSSAVNIMNGSQMHINPANKSLPPTFGPATLFNHFS SLFDSSQVPANQGWGDGPLSSRVATDASFTVQSAFLGNSVLGHLENMHPDNS KAPGFRPPSQRVSTSPVATSAPPTLGQPKGVSASQDRKIPPPIGTERLARIR QGGSSVAQAPAGTSFVAPVGHSGIWSFGVNAV
5819	A	1	394	SSLFDSSQVPAFQVRVSTSPVGLPSIDPSGSSPSSSSAPLASFSGIPGTRVFL QGPAPVGTSPFNHQHFSHPWTSASNSSTAPPTLGQPKGVSASQDRKIPPP IGTERLARIRQGGSSVAQAPAGTSFVAP
5820	A	2	1785	QLFACVPKTSPPATVISSVTSTCSSLPSVSSAPITSGQAPTTLFLPASTSQAQ LSSQKMFSAVPPTKEKVSTQDQPMANLCTPSSSTANSCSSASNTPGAPET HPSSSPTPTSSNTQEEAQPSVSDLSPMSMPFASNSEPAPLTLTSPRMVAAD NQDTSNLPQLAVPAPRVSHRMQPRGSFYSMVFNATIHQDPQSIFVTNPVTLT PPQGPAAVQLSSAVNIMNGSQMHINPANKSLPPTFGPATLFNHFS S LFDSS QVPANQGWGDGPLSSRVATDASFTVQSAFLGNSVLGHLENMHPDNSKAPGFR PPSQR\VSTSPVGLPSIDPSG\SSPSSSSAPL\ASFSGIPGNQGFLLQGP\A PVGGLLSFNQHF/SFPHPW\TSASNSCDSPIPSVSSGSSSPLSATS\APPT LG\QPKGSQCQSRIRKGYLPPIGTERLARILQGGSSVAQAPAG\TSFVAPVGH SG\IWSFGVNAV\EGL\SGWSQVMG\NHPMAFNFSGPKAHFSQHQPIMER DDSGMVAPSNIHFQPMASGFVDFSKGLPISMYGGTIIPSHPOLADVPGGPLF NGLHNPDPAWNPMIKVIQNSTECTDAQQVQWA
5821	A	3	125	HEKTAEPLKFPKRGPKPGSKVGKMWVSQKGIKIRDYKGVIR
5822	A	3448	3831	KNRFGSGVSSNSKSNNSCVYVYIDRDIDHTYIHIHTNICIHLFFFFFETE SHALSPRLECNGVISAHCNLHPPG\SSDSPASAARVAG\ITGTCHHAQLIFF FFVFLVETGFHHAQAQSQTDLR*STPLGFPC*DYRR/AAIVPGIFLLH* IR*\KVPTLLTDMRNASEYDCDFSTNKIDKEETFS*NASLNLCLLST*PYEM VTHFKGY*ILPLFFFFFLRQSL/SSVTQAGVQWHNLGSLQPLPPGFKQFSC SLPSSWDYRYQLPRLA/NF/FVFLVETGFHHAQAQSQTDLR

TABLE 21

TISSUE ORIGIN	RNA SOURCE	HYSEQ LIBRARY NAME	SEQ ID NOS:
adult brain	GIBCO	AB3001	82-83 127 137 157-159 179-181 217-219 287-292 356-359 460-461 474 507-508 559 577-578 674- 675 733-735 780-782 820-821 848-849 930 957 1005-1008 1011-1024 1034 1038-1039 1051 1059- 1063 1068-1074 1083-1084 1098 1142 1181 1186 1203 1214 1235 1265-1267 1339-1340 1363-1364 1373-1374 1402-1403 1412-1414 1451-1453 1455- 1457 1462-1470 1491 1497-1499 1514 1517-1519 1522-1524 1531-1533 1535 1566-1567 1576-1581 1609-1611 1625-1627 1644 1651-1653 1659-1664 1668 1702-1703 1707 1710-1711 1715 1726-1731 1774 1781 1789 1800-1808 1811 1838 1849-1851 1853 1857-1864 1876 1883 1885-1886 1909-1910 1941-1942 1963-1965 1975-1980 1982-1984 2013- 2014 2054-2057 2092 2156 2172-2175 2178-2185 2194-2196 2200 2202-2203 2216 2219 2234-2236 2262 2267-2268 2292 2320-2321 2361 2375-2377 2380-2382 2395-2397 2407-2408 2450-2451 2481- 2482 2486 2491-2494 2497-2502 2534 2547 2587 2622-2625 2627 2635 2642 2644-2646 2651 2661 2690 2703 2706 2717-2720 2765 2811 2854-2860 2885 2922 2989-2991 2994-2995 3024-3028 3033 3036 3096-3097 3112-3115 3120 3128-3129 3134- 3136 3139-3140 3167 3219-3221 3235-3236 3268 3396-3397 3399-3408 3427-3428 3430-3431 3434 3458 3467 3473-3474 3480-3482 3490 3500 3519- 3521 3538-3539 3542-3544 3551-3552 3567-3570 3586-3587 3608-3609 3662-3663 3725 3746-3747 3806-3813 3828-3832 3858 3964 3969-3971 4004- 4005 4012-4013 4047 4101-4104 4110-4112 4144- 4145 4171 4246-4248 4285 4339-4340 4345-4346 4384 4390 4404-4405 4415 4417 4428-4433 4524- 4525 4568-4569 4586-4590 4816 4843-4845 4866- 4868 4872-4873 4990-4991 5039-5040 5100-5101 5108-5109 5170-5171 5207 5214-5216 5460-5461 5593-5594 5618-5619 5621-5625 5787-5789
adult brain	GIBCO	ABD003	1-11 52 64 82-83 124 157-159 179-181 238 254 264-266 287-292 323-325 345 349 364 367-371 375 391-392 400-402 406 474 507-508 573 577- 578 590-593 598 614-615 624 674-675 693-694 701-702 706-708 743-747 768-769 820-821 825- 826 840-842 844 848-849 867-871 890-896 900- 902 905 913-917 920-928 941-946 948-952 957 960 963-969 977-989 991-994 998-1000 1003 1005-1008 1011-1024 1031-1032 1034 1038-1040 1049-1051 1059-1063 1065-1069 1083-1084 1089- 1090 1094-1098 1100-1101 1103-1104 1116-1121 1123-1124 1128 1146-1148 1167-1170 1174-1178 1186-1194 1203 1214 1216 1219 1222-1226 1234 1237 1242-1247 1261-1267 1282-1283 1300-1301 1309-1311 1314-1320 1339-1343 1351-1352 1360- 1362 1370 1373-1374 1380 1388-1389 1401-1404 1435-1436 1447-1448 1450-1453 1455-1456 1473 1491-1496 1501-1502 1514 1517-1520 1531-1533 1535 1543-1544 1549-1552 1557-1560 1562-1567 1574-1577 1584 1589-1590 1597 1613-1614 1619- 1620 1624-1627 1633-1634 1650 1654-1657 1659-

TABLE 21

			508 564-566 599-606 609-611 815-819 840-842
			848-849 963 990 1034 1040 1051 1100 1121 1203
			1265-1267 1270-1272 1295-1297 1370 1405-1409
			1449 1462-1470 1486-1487 1491 1517-1519 1534
			1549 1562-1565 1624 1669-1671 1704-1706 1714
			1743-1744 1752 1789 1814-1815 1817-1818 1849-
			1851 1876 1883 1885-1886 1901-1905 1928-1931
			1975-1980 2015 2058-2061 2092 2133 2168-2175
			2200 2205-2207 2217-2218 2292 2334 2373-2374
			2380-2382 2395-2397 2453 2472 2499-2500 2649
			2674 2765 2818 2892 2895 2932-2933 3018-3021
			3031 3053 3111 3116-3117 3214 3416-3417 3687
			3719-3723 3848-3851 3983 4004-4005 4024-4027
			4160 4217 4249-4251 4327-4329 4347-4349 4533-
			4534 4670-4671 4928-4931 4965 5100-5101 5335
			5460-5461 5495-5497 5712-5714 5809

TABLE 22

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	SMITH-WATERMAN SCORE	%IDENTITY
2	AJ243460	Leishmania major proteophosphoglycan	191	33.906
4	U42580	Paramecium bursaria Chlorella virus 1 PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	99	34.694
5	U42580	Paramecium bursaria Chlorella virus 1 PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	48	34.783
6	U42580	Paramecium bursaria Chlorella virus 1 PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	57	26.667
8	U42580	Paramecium bursaria Chlorella virus 1 PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	71	30.769
10	U42580	Paramecium bursaria Chlorella virus 1 PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	68	26.562
11	U42580	Paramecium bursaria Chlorella virus 1 PAPK (17X); similar to PBCV-1 ORF A41R, corresponds to Genbank Accession Number U17055	164	29.365
12	U80443	Caenorhabditis elegans contains similarity to a band 4.1-like domain; Pfam domain PF00373 (Band4_1) Score=132.4, E=8.4e-42, N=1	206	29.240
13	U80443	Caenorhabditis elegans contains similarity to a band 4.1-like domain; Pfam domain PF00373 (Band4_1) Score=132.4, E=8.4e-42, N=1	148	32.911
14	AF149422	Homo sapiens unknown	66	26.744
15	AF149422	Homo sapiens unknown	738	78.125
16	AF149422	Homo sapiens unknown	1229	50.117
17	AF149422	Homo sapiens unknown	76	40.541
18	AF149422	Homo sapiens unknown	58	36.111
19	AF149422	Homo sapiens unknown	60	30.769
20	AF149422	Homo sapiens unknown	102	82.353
21	AF149422	Homo sapiens unknown	67	20.588
22	AF149422	Homo sapiens unknown	66	20.571
23	AF149422	Homo sapiens unknown	4623	25.256
25	L24521	Homo sapiens transformation-related protein	98	50.000
26	L24521	Homo sapiens transformation-related protein	75	29.474
27	L24521	Homo sapiens transformation-related protein	314	71.765
28	L24521	Homo sapiens transformation-related	297	63.855

TABLE 22

5801	M22334	Homo sapiens unknown protein	2794	90.343
5802	M22334	Homo sapiens unknown protein	2108	92.793
5803	M22334	Homo sapiens unknown protein	2695	85.804
5804	M22334	Homo sapiens unknown protein	3731	92.707
5805	U87306	Rattus norvegicus transmembrane receptor UNC5H2	1213	91.753
5808	S58722	Homo sapiens X-linked retinopathy protein {C-terminal, clone XEH.8c}	143	86.957
5809	S58722	Homo sapiens X-linked retinopathy protein {C-terminal, clone XEH.8c}	143	86.957
5810	AL079292	Homo sapiens hypothetical protein, similar to (AC007017) putative RNA helicase A	4453	90.025
5811	AL079292	Homo sapiens hypothetical protein, similar to (AC007017) putative RNA helicase A	5080	98.767
5812	AL079292	Homo sapiens hypothetical protein, similar to (AC007017) putative RNA helicase A	253	97.297
5813	U48830	Mus musculus subtilisin-like proprotein convertase	163	59.184
5814	S58722	Homo sapiens X-linked retinopathy protein {C-terminal, clone XEH.8c}	311	58.333
5815	S58722	Homo sapiens X-linked retinopathy protein {C-terminal, clone XEH.8c}	237	74.545
5816	AB029008	Homo sapiens KIAA1085 protein	3242	86.897
5817	AB029008	Homo sapiens KIAA1085 protein	3769	96.558
5818	AB029008	Homo sapiens KIAA1085 protein	2424	83.691
5819	AB029008	Homo sapiens KIAA1085 protein	764	82.432
5820	AB029008	Homo sapiens KIAA1085 protein	3446	95.431
5822	AF010144	Homo sapiens neuronal thread protein AD7c-NTP	500	49.064

sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

5 The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a
10 virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be
15 confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each,
20 unless the context dictates otherwise.

5.2 NUCLEIC ACIDS AND PEPTIDES OF THE INVENTION

Sequences of the nucleic acids and peptides of the present invention are set forth in the Sequence Listing. Table 1 relates the SEQ ID's listed herein to their identification
25 in parent applications from which this application claims priority.

TABLE 1

Gene Family	Serial Number	Date Filed	SEQ ID NO. in Parent Applications	SEQ ID NO: in Current Application	File Name On CD
748	09/205,070	Dec. 3, 1998	SEQ ID NO: 1-45,196	748 SEQ ID NO: 1-45,196	N/A
	09/340,623	Jun. 28, 1999	SEQ ID NO: 1-45,196	748 SEQ ID NO: 1-45,196	

	09/898,888	Jul. 3, 2001	SEQ ID NO: 1-45,196	748 SEQ ID NO: 1-45,196	
752	09/205,155	Dec. 3, 1998	SEQ ID NO: 1-13192	752 SEQ ID NO: 1-13192	N/A
	09/359,922	Jul. 22, 1999	SEQ ID NO: 1-13192	752 SEQ ID NO: 1-13192	
	09/919,002	Jul. 30, 2001	SEQ ID NO: 1-13192	752 SEQ ID NO: 1-13192	
778	09/347,127	Jul. 2, 1999	SEQ ID NO: 1-94	778 SEQ ID NO: 1-94	N/A
	09/905,059	Jul. 12, 2001	SEQ ID NO: 1-94	778 SEQ ID NO: 1-94	
779	09/457,877	Dec. 8, 1999	SEQ ID NO: 1-128	779 SEQ ID NO: 1-128	N/A
	09/952,981	Sep. 14, 2001	SEQ ID NO: 1-128	779 SEQ ID NO: 1-128	
782	09/471,275	Dec. 23, 1999	SEQ ID NO: 1-10,451	782 SEQ ID NO: 1-10,451	Table2(782).doc
784	09/488,725	Jan. 21, 2000	SEQ ID NO: 1-10289	784 SEQ ID NO: 1-10289	Table3(784).doc; Table4(784).doc
	09/552,317	Apr. 25, 2000	SEQ ID NO: 1-10289	784 SEQ ID NO: 1-10289	
785	09/491,404	Jan. 25, 2000	SEQ ID NO: 1-3796	785 SEQ ID NO: 1-3796	(Table 5 and 7 are hard copies)
	09/922,279	Aug. 3, 2001	SEQ ID NO: 1-3796	785 SEQ ID NO: 1-3796	Table6(785).doc
787	09/496,914	Feb. 23, 2000	SEQ ID NO: 1-3960	787 SEQ ID NO: 1-3960	Table8(787).doc; Table9(787).doc;
	09/560,875	Apr. 27, 2000	SEQ ID NO: 1-3960	787 SEQ ID NO: 1-3960	Table10(787).doc
788	09/515,126	Feb. 28, 2000	SEQ ID NO: 1-14074	788 SEQ ID NO: 1-14074	Table11(788).doc; Table12(788).doc
	09/577,409	May. 18, 2000	SEQ ID NO: 1-14074	788 SEQ ID NO: 1-14074	Table13(788).doc
789	09/519,705	Mar. 7, 2000	SEQ ID NO: 1-6391	789 SEQ ID NO: 1-6391	Table14(789).doc; Table15(789).doc
	09/574,454	May. 19, 2000	SEQ ID NO: 1-6391	789 SEQ ID NO: 1-6391	Table16(789).doc
790	09/540,217	Mar. 31, 2000	SEQ ID NO: 1-30533	790 SEQ ID NO: 1-30533	Table17(790).doc; Table18(790).doc
	09/649,167	Aug. 23, 2000	SEQ ID NO: 1-30533	790 SEQ ID NO: 1-30533	Table19(790).doc
791	09/552,929	Apr. 18, 2000	SEQ ID NO: 1-5822	791 SEQ ID NO: 1-5822	Table20(791).doc; Table21(791).doc
	09/770,160	Jan. 26, 2001	SEQ ID NO: 1-5822	791 SEQ ID NO: 1-5822	Table22(791).doc
792	09/577,408	May. 18, 2000	SEQ ID NO: 1-8502	792 SEQ ID NO: 1-8502	(Table 23-25 are hard copies)

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of 748 SEQ ID NO: 1-45,196, 752 SEQ ID NO: 1-13,192, 778
5 SEQ ID NO: 1-94, 779 SEQ ID NO: 1-128, 782 SEQ ID NO: 1-10,451, 784 SEQ ID NO: 1-10,289, 785 SEQ ID NO: 1-3796, 787 SEQ ID NO: 1-3960, 788 SEQ ID NO: 1-14,074, 789 SEQ ID NO: 1-6391, 790 SEQ ID NO: 1-30,533, 791 SEQ ID NO: 1-5822, and 792 SEQ ID NO: 1-8502; a polynucleotide encoding any one of the peptide

6.0 EXAMPLES

Selected tables are provided on the CD-ROM, supplied concurrently with this application. The entirety of the contents of the CD-ROM are intended to be a part of this application, and are incorporated by reference. The electronic file name, where appropriate, follows the first reference to the table throughout the application.

6.1 The 748 Gene Family

Novel Contigs

The novel contigs of the invention, were assembled from novel expressed sequence tags (EST's) isolated by methods described herein (*e.g.*, SBH), and in some cases sequences obtained from one or more public databases. The inserts for the cDNA libraries from which the novel ESTs were obtained were amplified with PCR using primers specific for the vector sequences which flank the inserts. These samples were spotted onto nylon membranes and interrogated with oligonucleotide probes to give sequence signatures. The clones were clustered into groups of similar or identical sequences, and single representative clones were selected from each group for gel sequencing. The 5' sequence of the amplified inserts was then deduced using the reverse M13 sequencing primer in a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel ESTs. The novel contigs of the invention, were assembled from the novel ESTs and, in some cases, sequences obtained from one or more public databases. The sequences for the resulting contigs from the 748 gene family are designated as 748 SEQ ID NO: 1-45,196 and are provided in the Sequence Listing.

6.2 The 752 Gene Family

Novel Contigs

The novel contigs of the invention, were assembled from novel expressed sequence tags (EST's) isolated by methods described herein (*e.g.*, SBH), and in some cases sequences obtained from one or more public databases. The inserts for the cDNA libraries from which the novel ESTs were obtained were amplified with PCR using primers specific for the vector sequences which flank the inserts. These samples were

libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The inserts for the cDNA libraries from which the sequences were obtained were amplified with PCR using primers specific for the vector sequences which flank the inserts, or isolated from plasmid preparations. The 5' sequence of the amplified inserts was then deduced using the reverse M13 sequencing primer in a typical Sanger sequencing protocol, as well as internal primers in both the forward and reverse direction. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction. In all cases all of a signature cluster was sequenced to generate overlapping clones to assemble the contigs. Chromatograms were base called and assembled using a software suite from University of Washington, Seattle containing three applications designated PHRED, PHRAP, and CONSED. The sequences for the resulting contigs for the 782 gene family are designated as 782 SEQ ID NO: 1-10,451 and are provided in the attached Sequence Listing. inserts was then deduced in a typical Sanger sequencing protocol. The inserts of the library were, amplified with PCR using 5 primers specific for vector sequences which flank the inserts.

The contigs were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 114, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The nearest neighbor results for 782 SEQ ID NO: 1-10,451 are shown in Table 2, and identified as Table2(782).doc on the enclosed compact disc.

6.6 The 784 Gene Family

Novel Contigs

Table 3 (identified as Table3(784).doc on the enclosed CD) sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (784 SEQ ID NO: 1-10,289) of the present invention, and their corresponding nucleotide locations to each of 748 SEQ ID NO: 1-10,289. Table 3 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of translated novel polynucleotide to known polypeptides (W.R. Pearson, Methods in Enzymology, 183: 63-98 (1990), incorporated herein by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268: 78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame. When the predicted beginning nucleotide of Table 3 is a higher number than the predicted end nucleotide of Table 3, then the amino acid sequence is derived from the complementary strand of the indicated SEQ ID NO. The locations of the predicted beginning and end nucleotides correlate to the nucleotide sequence of the indicated SEQ ID NO., not its complementary strand.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in Table 3 or from six frame translations of 784 SEQ ID NO: 1-10,289; or the corresponding full length or mature protein. One of skill in the art could determine the corresponding amino acid sequence using techniques well known in the art to translate and analyze all possible six frames. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in the 784 SEQ ID NO: 1-10,289; or (b) polynucleotides that hybridize to the

Table 4 (identified as Table4(784).doc on the enclosed CD) shows the various tissue sources of the EST sequences from Hyseq's database which were used to assemble the contigs or nucleic acids of the present invention (identified by 784 SEQ ID NO: 1-10,289).

5 The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 114, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The
10 nearest neighbor results for 784 SEQ ID NO: 1-10,289 are shown in the Table 5, infra.

6.7 **The 785 Gene Family**

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared
15 from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide
20 probes (*e.g.*, 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to
25 fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

The novel contigs of the invention were assembled from sequences that were obtained from a cDNA library by methods described above, and in some cases sequences
30 obtained from one or more public databases. Chromatograms were base called and assembled using a software suite from University of Washington, Seattle containing three

applications designated PJRED, PHRAP, and CONSED. The sequences for the resulting contigs are designated as 785 SEQ ID NO: 1-3,796 and are provided in the Sequence Listing. The contigs were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (*i.e.*, Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 114, using Fastxy algorithm. Fastxy is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The nearest neighbor results for 785 SEQ ID NO: 1-3,796 are shown in Table 6 (identified as Table6(785).doc on the enclosed CD), *infra*.

The nucleotide sequence within the assembled contigs that codes for signal peptide sequences and their cleavage sites can be determined from using Neural network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrick Nielson, Jacob Englebrecht, Soren Brunak, and Gunnar von Heijne in the publication "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, vol. 10, no. 1, pp. 1-6 (1997) incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et. al., reference, are obtained from each assembled contig. Table 7 sets forth the nucleotide sequence range for each sequence of 785 SEQ ID NO: 1-3,796 that encodes a corresponding forty-five amino acid sequence containing the signal peptide sequence and its cleavage site, the maximum S score and the mean S score obtained for each sequence. Not all forty-five amino acids in the sequence may comprise the signal peptide.

6.8 The 787 Gene Family

Table 8 (identified as Table8(787).doc on the enclosed CD) sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (787 SEQ ID NO: 1-3960) of the present invention, and their corresponding nucleotide locations to each of 787 SEQ ID NO: 1-3960. Table 8 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of translated novel polynucleotide to known polypeptides (W.R. Pearson, Methods in Enzymology, 183: 63-98 (1990), incorporated herein by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268: 78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame. When the predicted beginning nucleotide of Table 8 is a higher number than the predicted end nucleotide of Table 8, then the amino acid sequence is derived from the complementary strand of the indicated SEQ ID NO. The locations of the predicted beginning and end nucleotides correlate to the nucleotide sequence of the indicated SEQ ID NO., not its complementary strand.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in Table 8 or from six frame translations of 787 SEQ ID NO: 1-3960; or the corresponding full length or mature protein. One of skill in the art could determine the corresponding amino acid sequence using techniques well known in the art to translate and analyze all possible six frames. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in the 787 SEQ ID NO: 1-3960; or (b) polynucleotides that hybridize to the complement of the

into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Table 9 (identified as Table9(787).doc on the enclosed CD) shows the various tissue sources of the EST sequences from Hyseq's database which were used to assemble the contigs or nucleic acids of the present invention (identified by 787 SEQ ID NO: 1-3960).

The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 114, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The nearest neighbor results for 787 SEQ ID NO: 1-3960 are shown in the Table 10, (identified as Table10(787).doc on the enclosed CD) infra.

6.9 The 788 Gene Family

Table 11 (identified as Table11(788).doc on the enclosed CD) sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (788 SEQ ID NO: 1-14,074) of the present invention, and their corresponding nucleotide locations to each of 788 SEQ ID NO: 1-14,074. Table 11 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of translated novel polynucleotide to known polypeptides (W.R. Pearson, Methods in Enzymology, 183: 63-98 (1990), incorporated herein by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268: 78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame. When the predicted beginning nucleotide of Table 11 is a higher number than the

Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

Novel Contigs

5 The novel contigs of the invention were assembled from sequences that were obtained from a cDNA library by methods described above, and in some cases sequences obtained from one or more public databases. The sequences for the resulting contigs are designated as 788 SEQ ID NO: 1-14,074 and are provided in the attached Sequence Listing; The contigs were assembled using an EST sequence as a seed. Then a recursive
10 algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (*i.e.*, Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences
15 into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

 Table 12 (identified as Table12(788).doc on the enclosed CD) shows the various tissue sources of the EST sequences from Hyseq's database which were used to assemble the contigs or nucleic acids of the present invention (identified by 788 SEQ ID NO: 1-
20 14,074).

 The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 115, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and
25 contains the translated amino acid sequences for which the assemblage encodes). The nearest neighbor results for 788 SEQ ID NO: 1-14,074 are shown in the Table 13, (identified as Table13(788).doc on the enclosed CD) infra.

6.10 The 789 Gene Family

30 Table 14 (identified as Table14(789).doc on the enclosed CD) sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (789

SEQ ID NO: 1-6,391) of the present invention, and their corresponding nucleotide locations to each of SEQ ID NO: 1-6,391. Table 14 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which
5 selects a polypeptide based on a comparison of translated novel polynucleotide to known polypeptides (W.R. Pearson, Methods in Enzymology, 183: 63-98 (1990), incorporated herein by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a
10 probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268: 78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest
15 open reading frame.

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using
20 primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (*e.g.*, 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a
25 typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

30 Novel Contigs

The novel contigs or the nucleic acids of the present invention of the invention were assembled from sequences that were obtained from a cDNA library by methods described above, and in some cases sequences obtained from one or more public databases. The sequences for the resulting contigs are designated as 789 SEQ ID NO: 1-6,391 and are provided in the attached Sequence Listing. The contigs were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (*i.e.*, Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UriGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95 %.

Table 15 (identified as Table15(789).doc on the enclosed CD) shows the various tissue sources of the EST sequences from Hyseq's database which were used to assemble the contigs or nucleic acids of the present invention (identified by 789 SEQ ID NO: 1-6,391).

The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 115, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The nearest neighbor results for 789 SEQ ID NO: 1-6,391 are shown in the Table 16, (identified as Table16(789).doc on the enclosed CD) *infra*.

6.11 The 790 Gene Family

Table 17 (identified as Table17(790).doc on the enclosed CD) sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (790 SEQ ID NO: 1-30,553) of the present invention, and their corresponding start and stop nucleotide location to each of 790 SEQ ID NO: 1-30,553. Table 17 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide

database containing EST sequences, dbEST version 115, gb pri 115, and UniGene version 10.3, and exons from public domain genomic sequences predicted by GenScan) that belong to this assemblage. The algorithm terminated when there were no additional sequences from the databases that will extend the assemblage. Further, the inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Table 18 (identified as Table18(790).doc on the enclosed CD) shows the various tissue sources of the EST sequences from Hyseq's database which were used to assemble the contigs or nucleic acids of the present invention (identified by 790 SEQ ID NO:1-30,553).

The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 1.15, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The nearest neighbor results for 790 SEQ ID NO: 1-30,553 are shown in the Table 19, (identified as Table19(790).doc on the enclosed CD) infra.

6.12 The 791 Gene Family

Table 20 (identified as Table20(791).doc on the enclosed CD) sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (791 SEQ ID NO: 1-5,822) of the present invention, and their corresponding nucleotide locations to each of 791 SEQ ID NO: 1-5,822. Table 20 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of translated novel polynucleotide to known polypeptides (W.R. Pearson, Methods in Enzymology, 183: 63-98 (1990), incorporated herein by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J.

Table 21 (identified as Table21(791).doc on the enclosed CD) shows the various tissue sources of the EST sequences from Hyseq's database which were used to assemble the contigs or nucleic acids of the present invention (identified by 791 SEQ ID NO: 1-5,822).

5 The nearest neighbor result for the assembled contig was obtained by a FASTA version 3 search against Genpept release 115, using FASTXY algorithm. FASTXY is an improved version of FASTA alignment which allows in-codon frame shifts. The nearest neighbor result showed the closest homologue for each assemblage from Genpept (and contains the translated amino acid sequences for which the assemblage encodes). The
10 nearest neighbor results for 791 SEQ ID NO: 1-5,822 are shown in the Table 22, (identified as Table22(791).doc on the enclosed CD) infra.

6.13 **The 792 Gene Family**

Table 23 sets forth the novel predicted polypeptides (including proteins) encoded
15 by the novel polynucleotides (792 SEQ ID NO: 1-8,502) of the present invention, and their corresponding nucleotide locations to each of 792 SEQ ID NO: 1-8,502. Table 23 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from <http://fasta.bioch.virginia.edu>) which selects a polypeptide based on a comparison of
20 translated novel polynucleotide to known polypeptides (W.R. Pearson, Methods in Enzymology, 183: 63-98 (1990), incorporated herein by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional
25 properties (C. Burge and S. Karlin, J. Mol. Biol., 268: 78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame.

30 The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in Table 23 or from six